

Psychological Bulletin

CONTENTS

ARTICLES:

The Freudian Theory of Daydreams: An Operational Analysis: WILLIAM SEEMAN, 369.

The History of the Leaderless Group Discussion Technique: H. L. ANSBACHER, 383.

STATISTICS:

Computation of the Level of Significance in the F-Test: C. J. BURKE, 392.

On the Use of Latin Squares in Psychology: QUINN MCNEMAR, 398.

The Genetics of Schizophrenia: Reply to Pastore: LEWIS A. HURST, 402.

Correlation versus Curve Fitting in Research on Accident Proneness: Reply to Maritz: MILTON L. BLUM AND ALEXANDER MINTZ, 413.

Concerning Taste-Blindness to PTC: JOSEF COHEN AND DONALD P. OGDON, 419.

Reply to Travers' "A Critical Review of the Validity and Rationale of the Forced-Choice Technique": DONALD E. BAIER, 424.

Reply to Travers' Critical Review of the Forced-Choice Technique: MARION W. RICHARDSON, 435.

REVIEWS:

Mowrer's Learning theory and personality dynamics: E. R. HILGARD AND E. L. WALKER, 438.

Lewy's Hypnosetherapie und psychosomatische Probleme: HANS H. STRUPP, 442.

Remblueth's The transmission of nerve impulses at neuro-effector junctions and peripheral synapses: LORRIN A. RIGGS, 444.

Childner's Studies in leadership: LAURIE F. CARTER, 445.

Entail's Public opinion 1933-1946: ROSE SEAMER, 447.

Wirt's Selected readings in social psychology: KIMBALL YOUNG, 448.

Lehman's The human group: M. BREWSTER SMITH, 449.

Burn's Practical and theoretical aspects of psychoanalysis: JESSIE L. MILLER AND JAMES G. MILLER, 451.

Mowrer's The meaning and practice of psychotherapy: O. H. MOWRER, 453.

Waring's Some theory of sampling: FREDERICK MOSTELLER, 453.

Myer and Henry's Handbook of applied psychology: ROBERT L. THORNDIKE, 453.

Whitfield's The new you and heredity: LEWIS M. TERMAN, 456.

Wickens's Theory of mental tests: QUINN MCNEMAR, 458.

AND MATERIALS RECEIVED: 461

LYLE H. LANIER, Editor
University of Illinois

MARGARET K. HARLOW, Managing Editor

CONSULTING EDITORS

STUART H. BRITT
McCann-Erickson, Inc., New York

DORWIN CARTWRIGHT
University of Michigan

FRANK A. GELDARD
University of Virginia

JAMES J. GIBSON
Cornell University

DAVID A. GRANT
University of Wisconsin

WILLIAM T. HERON
University of Minnesota

ERNEST R. HILGARD
Stanford University

WILLIAM A. HUNT
Northwestern University

JEAN WALKER MACFARLANE
University of California

DONALD G. MARQUIS
University of Michigan

JOHN T. METCALF
University of Vermont

JAMES G. MILLER
University of Chicago

NEAL E. MILLER
Yale University

HELEN PRAK
University of Michigan

ROBERT R. SEARS
Harvard University

ROBERT L. THORNDIKE
Teachers College, Columbia University

The Psychological Bulletin contains evaluative reviews of the literature in various fields of psychology, methodological articles, critical notes, and book reviews. This JOURNAL does not publish reports of original research or original theoretical articles.

Editorial communications, manuscripts and book reviews should be sent to Lyle H. Lanier, Department of Psychology, University of Illinois, Urbana, Illinois.

Preparation of articles for publication. Authors are strongly advised to follow the general directions given in the article by Anderson and Valentine, "The preparation of articles for publication in the journals of the American Psychological Association" (*Psychological Bulletin*, 1944, 41, 345-376). Special attention should be given to the section on the preparation of the bibliography (pp. 363-372), since this is a particular source of difficulty in long reviews of research literature. *All copy must be double-spaced, including the bibliography.*

Reprints. Fifty reprints are given, if requested, to contributors of articles, notes and special reviews. Five copies of the JOURNAL are supplied gratis to the authors of book reviews.

Business communications—including subscriptions, orders of back issues and changes of address—should be sent to the American Psychological Association, 1515 Massachusetts Avenue, N. W., Washington 5, D. C.

Annual subscription: \$7.00 (Foreign \$7.50). Single copies, \$1.25.

PUBLISHED BI-MONTHLY BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

1515 Massachusetts Ave., N.W., Washington 5, D.C.

Entered as second class mail matter at the post office at Washington, D.C., under the act of March 3, 1879. Additional entry at the post office at Menasha, Wisconsin. Acceptance for mailing at special rate of postage provided for in Section 538, act of February 25, 1925, authorized August 6, 1945.

Copyright, 1951, by The American Psychological Association, Inc.

Psychological Bulletin

THE FREUDIAN THEORY OF DAYDREAMS: AN OPERATIONAL ANALYSIS¹

WILLIAM SEEMAN
University of Pennsylvania

INTRODUCTION

The formulation of psychodynamics of daydreaming by Sigmund Freud (9, 10, 11, 13) enjoys a distinctive and interesting status in the domain of behavior science which is demarcated as "clinical"; for, following Freud, it is almost universally accepted in clinical circles and in the clinical literature that daydreaming serves a gratification function—i.e., that it is wish-fulfilling in character. Indeed, so strong is this conviction with respect to the psychodynamic function of the daydream that in at least one instance (41) it is incorporated into the *definition* of daydreaming, which is said to be "the imaginary representation of satisfactions that are not attained in real experience" (41, p. 186). Yet, notwithstanding the strength of this conviction and the widespread acceptance of the theory, it does not appear that it has ever been empirically demonstrated. Indeed, there is apparently no formal or operational analysis in the literature which indicates *how* the theory may be laid open to empirical investigation, and consequently to verification—to confirmation or disconfirmation. Such an analysis is nevertheless an indispensable prerequisite to empirical investigation. As it stands, the theoretical formulation is inadequate in operational specifi-

¹ This paper is adapted from a thesis presented to the Psychology Department of the University of Minnesota in partial fulfillment of the Ph.D. degree. I am indebted to Dr. Laurence F. Shaffer, of Teachers College, Columbia University, for time given and for his generosity in permitting me to see some of his unpublished daydream data; to Dr. Wallace A. Russell, whose incisive comments and criticism aided materially in clarifying certain formal aspects of the analysis; to my brothers, Dr. Julius Seeman, of the University of Chicago, and Dr. Melvin Seeman, of the Ohio State University; to Dr. Kenneth MacCorquodale, who in the course of three years of contact has contributed in no small measure to my thinking on this problem; and, finally, to Dr. Paul E. Meehl, whose encouragement and support were of inestimable value in the execution of the thesis investigation.

TABLE I

SUMMARY OF LITERATURE RELEVANT TO THE FREUDIAN THEORY OF DAYDREAMS

Author*	Subject	Methodology	Quantification	Conclusions
Bose (2)	Use of daydreams for therapy in psychoanalysis	Report and interpretation of use of fantasy with patients	None	Asks patient to "indulge in such daydreams as would give imaginary satisfaction to those [i.e., repressed] wishes" (2, p. 34)
Conklin (3)	Foster child fantasy	Questionnaire	Percentages	Foster child fantasy a "common" experience; reported by 24% of males, 31% of females
Conklin (4)	Relation between daydreaming and test adjustment	Personality inventory (Thurstone)	Percentages†	Relation between "neurotic" scores and "frequent" daydreaming
Dexter (5)	Imagination	Projective and structured personality tests; exciting events; sentence completions	Correlation	Correlations low positive most of them not significant; highest <i>r</i> (Bernreuter with self-rating) .527
Eidelberg (6)	Masturbation fantasy	Psychoanalytic interpretation of fantasy	None	Progressive changes in fantasy provide barometer for measuring therapeutic progress; may shorten therapy if understood
Ferenczi (7)	Gulliver fantasies	Psychoanalytic interpretations of <i>Gulliver's Travels</i>	None	"Compensatory wish-fulfilment" (7, p. 286) where fantasy involves size reduction
Fülop (17)	Relation of fantasy to body structure	Analysis of literature and clinical data	None	Fantasy life helpful or harmful, depending on morphology
Green (18)	Daydreams of child and adult clinical cases	Intuitive clinical analysis of case material	None	Accepts Freud's formulation but uses McDougall's concepts (e.g., daydreams gratify "gregarious instinct," and others (18, p. 19)
Hart (19)	Variety of fantasy	Interpretation of daydream material	None	Daydreams viewed as "building of pleasant mental pictures in which the complexes attain an imaginary fulfilment" (19, p. 155)
Hesnard (20)	Erotic daydreams	Psychoanalytic interpretation of sexual fantasy	None	Function of daydream: "Procurer... des excitations érotiques... plus ou moins intentionnelles." (20, p. 524)
Hurlock and Burnstein (24)	Imaginary playmate	Questionnaire	Percentages	Imaginary playmate reported by 31% of females, 23% of males; 80% of these had strong positive affect; no instances of strong negative affect
Jaehner (25)	Imaginary companion	Interpretation of child fantasy	None	Inferiority feelings generate compensatory fantasy
Kamiat (26)	Cosmic fantasy	Speculative analysis of cosmic fantasy	None	Cosmic fantasy compensates for feelings of insecurity and inadequacy
Lehrman (29)	Variety of daydreams	Analysis of clinically observed behavior	None	Daydreams are wish-fulfilling; foster child fantasy "a compromise between incestuous wishes and the neurotic flight from it" (29, p. 343)
Lévy-Valensi (30)	"Bovarysmes et constitutions mentales"	Analysis of mental activity analogous to that of Madame Bovary	None	"Primitive" bovarysm not wish-fulfilling; "secondary" bovarysm "s'offre comme une compensation, une fuite dans la fiction pour échapper au réel" (30, p. 295)
Rosenzweig (38)	Methods of studying fantasy	Didactic discussion of difficulties and methods	None	"Infantility" of daydreaming stressed; child's method of securing "immediate satisfaction despite adverse circumstances" (38, p. 42)

TABLE I (Continued)

Author*	Subject	Methodology	Quantification	Conclusions
Ruenauer (40)	Variety of day-dreams	Questionnaire; day-dream descriptions of 65 Indian students	Percentages	Grandeur, personal exploits, and political freedom daydreams reported; some sex differences noted
Shaffer (41)	Variety of day-dreams	Questionnaire	Percentages, means, and sigmas	Daydreams a mechanism of defense and source of satisfaction for the normal individual as well as the maladjusted
Smith (43)	Variety of day-dreams	Description of day-dream types	None	"Mental device for encompassing all desires"; daydreams relevant to needs; vary with sex and socio-economic status
Varendonck (48)	Fantasy activity of author	Introspective analysis of genesis, content, termination of own thought chains	None	Daydreams "practically a search for pleasurable representations" (48, p. 15); differs from "logical ideation, which corresponds to reality" (48, p. 13); "directed by one or several wishes" (48, p. 276)
Van Waters (47)	Imaginary companion of a little girl	Analysis of fantasy and clinical material	None	Fantasy activity regarded as providing substitute gratification
Vostrovsky (49)	Imaginary companions	Questionnaire	Percentages	Imaginary companion may appear as late as adolescence
Woodworth (53)	General nature of fantasy	Analysis of day-dream types appearing in literature	None	Daydream function: "getting for the moment the satisfaction of some desire" (53, p. 494)

* Listed in alphabetical order.

† Although Conklin reports only percentages and does not submit his data to tests of significance, he provides enough information to permit this. The computed χ^2 , disregarding the contribution of one cell which fails to meet the criterion of theoretical $N=5$, is 59.19. This is significant beyond the 1 per cent level.

‡ Defined as "le pouvoir à l'homme de se concevoir autre qu'il n'est" (30, p. 289).

cations which would make it possible to derive such hypotheses as might lend themselves to quantitative confirmation within specified probability levels.

Some idea of the extensive acceptance of Freud's theory of daydreaming may be indicated by Table I, which presents in summary form a review of a good deal of the literature relevant to the theory. The column marked "Conclusions" indicates the frequency with which the gratification function assigned to the daydream by Freud is accepted by writers on the subject. One might get the impression that many clinicians suppose the theory to be rather solidly grounded in empirical evidence. Yet Conklin's lament that "of systematic studies we have not so many" (4, p. 217) may be echoed today, fifteen years later. As the table suggests, there have been few attempts at quantification: only one of the studies (5) employs correlation techniques; five (3, 4, 24, 40, 41) report percentages, and of these only Conklin's (4) and Shaffer's (41) data permit the calculation of tests of significance.

A critical evaluation of the comparative merits of these writings would require a careful and explicit distinction between the fruitfulness of *content* and methodological soundness and rigor. It should be clear

that a formulation may be rich on the content side (i.e., potentially contain many hypotheses), but at the same time may suffer from severe methodological weaknesses. This would appear to be the case with Freud's formulations, which provide insightful hypotheses of a remarkably penetrating caliber for empirical investigation, and which have, as Mowrer and Ullman (35) have noted, considerable pragmatic merit. With respect to most of the other psychoanalytic papers cited (2, 6, 7, 20, 26), however, it would appear that they contribute little that is new, and that they are notably deficient in recognition of the formal problems involved.

Conklin's earlier paper (3) provides objective support for the psychoanalytic contention that the foster child fantasy is a common one; but in view of his earlier citation of Karl Abraham to the effect that the fantasy is universal, he appears to strain his own evidence considerably in stating that "Comparison with the psychoanalytic presentation . . . results in both support and amplification of the generalization from psychoanalysis" (3, p. 20). His college study (4) is a significant contribution to the problem in that it demonstrates statistically highly reliable covariation of personality inventory scores with frequency of reported daydreams, the covariation being in the direction which the theory would appear to require. Shaffer's study (41) likewise provides evidence which supports Freud's theory. The data confirm Freud's assertion as to the normality of daydreaming and justify Shaffer's statement that "daydreaming is an exceedingly common and therefore, in the statistical sense, a normal form of behavior" (41, p. 195). They also have deeper theoretical significance in that they demonstrate quantitative differences which are consistent with the formulation.

Freud laid out the general psychodynamic nature of daydreaming in several of his writings. In 1900 he stated:

A more thorough examination of the character of these day-phantasies shows with what good reason the same name has been given to these formulations as to the products of nocturnal thought—*dreams*. They have essential features in common with nocturnal dreams; indeed, the investigation of daydreams might really have afforded the shortest and best approach to the understanding of nocturnal dreams. Like dreams, they are wish-fulfilments . . . (13, p. 457).

Later, in 1908, in a paper on hysterical fantasies, Freud wrote:

These phantasies are wish-fulfilments, products of frustration and desire; they are justly called daydreams, for they give us a key to the understanding of night dreams, the nucleus of which is nothing else than those daytime phantasies, but complicated and distorted (11, p. 52).

And in the same year, writing on the poet and daydreaming:

Let us try to learn some of the characteristics of daydreaming. We can begin by saying that happy people never make phantasies, only unsatisfied ones. Unsatisfied wishes are the driving power behind phantasies; every separate phantasy contains the fulfilment (15, p. 176).

Again, in his *General Introduction*, Freud writes, "... daydreaming also is a mode of activity closely linked up with gratification, which is, in fact, the only reason why people practice it" (10, p. 117); and elsewhere in the same volume, "Now daydreams are literally wish-fulfilments . . ." (10, p. 117).

Although, as White (51) points out, some rather formidable problems compelled Freud in 1933 to state that dreams were *attempted* wish-fulfillments which, under certain circumstances, could achieve their ends only incompletely, there is some reason to doubt that he ever really abandoned the original form of his theory. For later, in 1936, we find him again writing:

... but the isolated thought is found to be an impulse in the form of a wish, often of a very repellent kind, which is foreign to the waking life of the dreamer and is consequently disavowed by him with surprise and indignation. This impulse is the actual constructor of the dream . . . (8, p. 80).

It is not here contended that this view of the psychodynamics of the daydream was wholly original with Freud, nor was it his exclusive property. In fact, Freud himself (11) credits Havelock Ellis with it, and cites Breuer, Janet, and Pick as well. We have seen, too, how widely accepted it is in the clinical literature and in clinical circles, irrespective of other theoretical commitments. It is true, however, that the formulation is most systematically made by Freud, who, in fact, goes on to discuss its similarity to nocturnal dreaming and to the formation of neurotic symptoms. It becomes an integral part of his theoretical structure. An analogous situation may be found with respect to Postulate 9 in Hull's (23) system, a postulate on conditioned inhibition which is taken almost intact from Pavlov and incorporated as a part of Hull's theory.² In fact, a large part of Hull's formulations have an honored history in psychology; and this is only what one would expect in the biography of a science. This obtains no less in physics, where, prior to Einstein's formulations of relativity theory, other distinguished efforts were made to solve some of the paradoxical results of experimentation.

LEVELS OF THEORY

A scientific theoretical structure may be viewed as a language system; and, if the term "primitive" is used in the same sense in which

² In his most recent statement (21), this is Postulate X.

it is used by Whitehead and Russell in *Principia Mathematica* (52), a relation of "more primitive" may exist between two language systems. The term "levels" is used here to convey the notion of this kind of relationship. In psychology the distinction between levels of analysis has been made in terms of molar versus molecular levels (23, 44, 45). The same set of behavioral events, as Spence (44) has pointed out, may be described in languages which are at different levels of analysis. In behavior theory this has apparently led, at times, to the notion that the two descriptions are at variance. About this, Spence writes:

Such different descriptions, however, do not necessarily represent fundamental disagreements. If the two systems of concepts should each be successful in leading to the discovery and formulation of laws, it should also be possible to discover co-ordinating definitions which will reveal the interrelations of the two systems. Or, as Hull suggests, the postulates . . . at a more molar level may ultimately appear as theorems in a more molecular description (44, p. 71).

A further illustration of the concept of levels of theory is provided by Hull (22), and here we note that the distinction in levels is not made along molar-molecular lines. After elaborating a miniature system of adaptive behavior, consisting of eighteen definitions and six postulates, Hull develops a number of theorems. In Theorem XII he states: "*Organisms capable of acquiring anticipatory goal reactions will strive to bring about situations which are reinforcing.*" Such "striving" to bring about a goal state of affairs will be recognized at once as fitting the concept of "motivation," and Hull therefore states in a footnote: "An additional element of interest in this theorem is the fact that the fundamental phenomenon of motivation seems to have been derived from the ordinary principle of association . . ." (22, p. 14). Thus in the language system of learning theory the concept of motivation is regarded as derivable "from associationist principles,"³ in contrast to its status at the "clinical" level, where motivation is taken as a "primitive" concept. Such distinction between the "learning" and "clinical" languages is involved in discussions by Mowrer (32), Miller (31), Mowrer and Lamoreau (34), Mowrer and Ullman (35), and Mowrer and Kluckhohn (33). These writers also attempt to discover relationships between the two language systems.

The attention which clinical psychologists have given to serious theoretical formulation has been, until recently, only moderate in degree. And it is perhaps this, as much as anything else, which accounts for the fact that clinical psychology has developed almost a separate

³ It is not intended here to convey the notion that this has, in fact, been successfully effected. On this, see Koch (27).

biography as a behavior science, and that there has been minimal communication between systematic (academic) psychology and clinical psychology. Doubtless the intensely pragmatic demands of clinical psychology as an "applied" domain also account considerably for this failure of communication. Recently, however, clinical psychologists have been demonstrating interest in systematic formulations, and the possibility is even envisioned that clinical psychology might become "in considerable measure the content of systematic psychology" (39, p. 5). To anyone who regards it as desirable that psychology eventually develop the kind of formal and systematic structure which is perhaps a defining property of scientific maturity, this turning of attention on the part of clinical psychologists to problems of theory and of theory building at the "clinical" level must be regarded as a salutary development. And this is true not only because a body of coherent theoretical formulations has heuristic value in sharpening problems and indicating avenues and areas of investigation, but also because it compels a consideration of how investigation and theory in the clinical area can most fruitfully be incorporated into a unified discipline of behavioral science.

Already there are a number of formal structures at the clinical level: a provocative and intriguing attempt at theory building at this level has been made by Murray *et al.* (36). In their well-known *Explorations in Personality* they have elaborated a system of theoretical constructs including need, press, thema, regnancy, and others. The system is frankly centralistic in character, has a certain amount of imaginative appeal, and constituted at the time of its publication a daring departure from the strict criteria of a narrowly conceived behaviorism. It remains, however, quite incomplete. Another attempt at theory building at the clinical level is that of Rogers, who has called his mimeographed form a "tentative draft" (37). The "theory" consists of a series of seventeen propositions, each elaborated at some length and dealing with: (a) the reaction of the organism as "an organized whole"; (b) the nature of his perceptual behavior; (c) the concept of the self; (d) the nature of psychological maladjustment, which is held to be a consequence of "a discrepancy between the organic perceptions and the self-concept" (37, p. 13); and (e) the effect of threat on the self-concept. Some attempt at systematic formulation is also to be found in Lecky's (28) little book, where the concept of self-consistency is central to the "theory."

The only formulation at the clinical level which, in my opinion, may be properly designated as a theory of behavior is that of psychoanalysis. This statement is not intended to convey any notion with respect to

the "truth" value of psychoanalytic theory. It has reference to the comprehensiveness of the theoretical structure and to the relations which exist among the propositions. It is recognized that a theory may be most comprehensive and at the same time "false" in the sense that (a) it asserts propositions which are contradicted by empirical and experimental evidence, and (b) it asserts propositions which are mutually contradictory. Unfortunately, the most cogent statement of the system in rigorous terms has not yet been made, and some of its adherents and practitioners admit that much psychoanalytic writing is fluid, ambiguous, and unintegrated. To repeat, therefore: there is no intent here to assert that psychoanalysis is the only "true" system of psychodynamics, whatever that might mean. The reference is merely to the fact that Freud appears to have been the only theorist who has laid out with some comprehensiveness and care an explanation on the clinical level as to the nature of human behavior.

OPERATIONAL ANALYSIS

An integral proposition in this framework of psychoanalytic theory is that which assigns to the daydream the general psychodynamic function of wish-fulfillment. It will be my purpose in the remainder of this paper to present a formal analysis of the problems presented in an empirical verification of this theory and to suggest a "methodological model" for an investigation. It appears to be the primary objective of such an analysis to indicate the defining operations which are indispensable to casting the theory into a form susceptible of quantitative empirical investigation.

It seems appropriate, in connection with this problem, to consider first the significance of the *wish language*.⁴ If one sticks close to Freud's intent, it would appear that the definition of a wish should be in terms of goal objects and/or goal states of affairs. While Freud's "Wünsche" could probably be developed in terms of some *drive* language, and while such a development could be justified on pragmatic grounds if it turned out to be more fruitful, that would still constitute some departure from what Freud appears to have in mind in his use of the wish language.⁵

⁴ The choice of the word "wish" in the translation has more than trivial consequences, and it is therefore important to note that it is an accurate rendering. In the original German we find "Diese Phantasien sind Wunschbefriedigungen . . ." (12, p. 192); and "Unbefriedigte Wünsche sind die Triebkräfte der Phantasien . . ." (9, p. 216). The significance of this lies in the implications of the wish language which will be developed.

⁵ The use of a drive terminology would present other problems as well. For one thing, as Skinner (42) points out, there is no *one* conceptual drive formulation in psychology which is accepted by all psychologists. In some instances it is regarded as a *stimulus*

Such a departure would appear to be unnecessary, for Tolman (45) has developed a concept, the concept of "demand," which may be defined in terms of goal objects and/or goal states of affairs. In some detail, he presents experimental operations "indicating the reality and objective definition of the rat's 'demand for' specific types of goal objects" (45, p. 37). As developed in his book, however, the concept is somewhat limited for the purpose of the present task. It needs to be expanded to include more complex states of affairs—e.g., "demand for vocational success," "demand for personal attractiveness," demands for sex and food objects. Such expansion presents no major formal problems. The defining operation for any demand (wish) can be explicitly described in terms of verbalizations of a specified kind; e.g., a "demand for success" might be defined in terms of verbalizations with respect to the acquisition of certain kinds of objects. But there is no logical reason to assume that the defining operations must be *limited* to verbal behavior. We may at the present time leave it an open question whether a concept like "unconscious demand" ("unconscious wish") will be necessary, or whether a defining operation can be found for such a concept.

Given specific defining operations for specific types of wishes (demands)—such as demand for a sex object, for "vocational success," for "physical attractiveness," etc.—the Freudian theory that daydreams are wish-fulfillments may be restated as follows: *The emission of a daydream is functionally related to a specific type of demand (wish), the relation being such that whenever an instance of such and such a daydream is observed, it is required by the theory that an instance of a specified corresponding demand (wish) must be identified by a suitable objective operation.* It seems clear that, so stated, the theory really requires the occurrence of identifiable, lawful patterns of *demand-daydream* covariation. What is crucially important here is the understanding of the *contingent notion of frequency*, which lies buried in this analysis of the *meaning* of the concept of wish-fulfillment. A concept of "frequency" is involved in the *definition* in the sense that there is implied the notion that whenever an instance of a daydream is counted an instance of a relevant demand must be identified. That this is *not* an assumption apart from the *definition* of wish-fulfillment, but a contingent condition which it imposes, is the essential point at issue here. The significance

(41); this would pose great difficulties. In others (21), the stimulus properties are predominant, though not exclusive. For most psychologists its physiological correlates appear to be crucial. Finally, the use of a drive terminology might conceivably pose formal problems in an attempt to distinguish between primary and secondary drives.

of this point lies in the fact that specific deductions from the theory will be essentially *predictions of where these frequencies may be expected to lie in consequence of theoretical requirements*.

It is necessary also to introduce explicitly the concept of "stronger than" as a relation between demands. This can be done by definition, and in any one of several alternative ways, each of which would be adequate to the purposes of the present task. Tolman's own definition, as introduced in the following passage, is quite acceptable:

Turning now to the results, it appears that the groups may be arranged in the order of goodness of their performances. It appears, in short, . . . certain goal objects or situations produce better total maze performance than do others. And this introduces us to the conception that certain goal objects are . . . demanded more than others The strength of the demand for the type of goal object provided is thus one of the immediate immanent aspects inherent in, and defining itself through, maze performances (45, p. 41).

This defining operation has certain similarities to Warden's methodology (50), in which the drive-measuring operation is likewise essentially a rank-ordering with respect to performances. However, there is available a convenient alternative procedure, more nearly isomorphic with the defining operations for "denser than" cited by Bergmann and Spence (1, p. 8) in which liquid Y is said by definition to be denser than liquid X if X floats on Y. Such an operation has already been performed by Tsai (46) in comparing the relative strength of sex and hunger motives. The analogous operation with respect to the present problem would be as follows: where it has been empirically demonstrated that a choice is made at the 1 per cent level of confidence between two objects or states of affairs, the relation of "stronger than" will be said by definition to hold for the specified demands.⁶

The formal aspects of an investigation of this character, which aims at empirical verification of a theory, ordinarily takes the form of an inductive leap from $(P \supset Q) \cdot Q$ to (P) where P is the theoretical formulation (in this instance "daydreams are wish-fulfillments") and Q a class of observable behavioral events (e.g., responses).⁷ However, it

⁶ In selecting such a procedure as a defining operation, however, it must not be forgotten that there is no *logical* reason to assume a transitivity relationship. That is, empirically, it is at least possible that A can be in greater demand than B and B in greater demand than C without assurance that A will be in greater demand than C. Should this turn out to be the case, the defining operation will have proved inadequate, and it would then be necessary to introduce the concept in a manner more nearly isomorphic with the Tolman and the Warden procedures.

⁷ This is, of course, an extremely elementary application of symbolic logic. The symbol " \supset " is the symbol of *implication*, and the expression $P \supset Q$ is read "if P, then Q," or "P implies Q." The symbol " \cdot " indicates a "binding" operation, and is usually

frequently happens that a proposition or predicate (Q) is entailed not by a single proposition but by several, and the symbolic statement in the case of the problem under consideration would take the form

$$[(P \cdot P_1' \dots_k) \supset Q_1 \dots_k] \cdot [Q]$$

where P represents the major proposition (i.e., the theory) under investigation and $P_1' \dots_k$ represents in each instance either a known empirical fact about the behavior of organisms (e.g., a known fact about adolescent interests and preferences) or an assumption.⁸ The several propositions, then, P and P' , entail a consequence, a predicate, Q ; and it is this logically derived consequence which will be, in each case, the predicate which constitutes the hypothesis (H) which the empirical data would be required to confirm or unconfirm. In instances where the confirmation conditions show Q to be in fact the case, there is nothing embarrassing about this; but where the confirmation conditions fail to do so, and $\sim Q$ is the case, then there are two sources of possible error, and it would be difficult to determine whether the source is P or P' . The argument here would be, however, that if in a large number of instances (i.e., test conditions) involving the same P but different P' 's there is overwhelming confirmation of Q , then in those isolated instances (should such occur) where the confirmation conditions indicate that $\sim Q$ is the case, the indication would be for a re-examination of P' before P .

In conclusion I should like to illustrate the procedure under this "methodological model." This may be done by examining some of the data presented by Ruenaufuer (40), already cited in Table I. He informs us that there is a measurably different intensity between male and female college students in India with respect to the daydream of independence. It could undoubtedly be established by objective operational procedures that the demand (wish) for independence is more "masculine" in the sense that it is so rated by *both* males and females in India.

read as "and." Whether because of the advantages of clarity and precision of deduction, or because of other advantages, symbolic logicians have apparently extended the power of analysis beyond that which has been characteristic of the classical logic. Those interested in a more complete description of the nature of symbolic logic are referred to Whitehead and Russell (52).

⁸ An alternative way of stating this would be that P' represents a class of propositions which constitute statements about known empirical facts or about assumptions. Since, in the event of the occurrence of confirmation conditions for any given Q , both P and P' are confirmed, there is no formal difference between the "theory" and the "assumption." However, the theory is the proposition P , which occurs in each and every hypothesis unchanged, whereas the assumption proposition is different for each H .

The actual technical procedures (i.e., the nature of such a rating operation) need not concern us in this exemplification of the formal aspects of the investigation. We would then have the following:

P_1' , i.e., the demand (wish) for independence is differentially stronger in Indian college males; (P_1')

P_2' , i.e., the intensity ratings for daydreams are a function of differential demand (wish) strength; (P_2')

HYPOTHESIS: $P \cdot P_1' \cdot P_2' \supset Q_1$ (H_1)

where Q_1 is a proposition asserting significant difference in intensity ratings; that is, if it is the case that daydreams are demand-relevant (i.e., wish-fulfilling), and in consequence of the known differential demand strength already established, and in further consequence of the functional relation between demand strength and intensity rating, it is theoretically required that the college males in the specified sample report significantly greater intensity ratings than do females.

The same formal procedures would lead to hypotheses about behavioral events which are not yet known, or at least not reported by Ruenauffer. If we consider the *frequency* with which this daydream type is reported, we should state as an hypothesis the following:

HYPOTHESIS: $P \cdot P_1' \supset Q_1'$ (H_2)

that is, if it is the case that daydreams are demand-relevant (i.e., wish-fulfilling), and in consequence of the known differential demand strength, it is theoretically required that this daydream type be experienced with a significantly greater frequency by the male Indian college students.

While the confirmation conditions for this hypothesis are unfortunately not reported by Ruenauffer, there is some hint that the original data might well confirm the hypothesis in his statement that "even" 50 per cent of the females report it, for this certainly suggests that a good many *more* of the males did.

SUMMARY

This paper attempts an operational analysis of Freud's theory of daydreams, and it is contended that such an analysis is an indispensable prerequisite to any investigation aimed at empirical confirmation of the theory. The concept of the wish is defined in terms of an expanded version of the operation defining Tolman's concept of "demand." Formally, the procedure is hypothetical-deductive in the sense that specific hypotheses for which confirmation or unconfirmation in any

investigation would be sought, are derivable from the theory, together with certain other propositions. Like all such investigations of an empirical character, however, any investigation of the theory would take the form of an inductive leap from $(P \supset Q) \cdot Q$ to (P) .

The paper also discusses "levels" of theory and makes a distinction in levels between the "clinical" and "learning" language systems, with special reference to the problem of motivation.

BIBLIOGRAPHY

1. BERGMANN, G., & SPENCE, K. W. The logic of psychophysical measurement. *Psychol. Rev.*, 1944, 51, 1-24.
2. BOSE, G. Opposite fantasies in the release of repression. *Indian J. Psychol.*, 1935, 10, 29-41.
3. CONKLIN, E. S. Foster child fantasy. *Univ. Oregon Publ.*, 1920, 1, No. 4.
4. CONKLIN, E. S. *Principles of adolescent psychology*. New York: Holt, 1935.
5. DEXTER, EMILY. Relation of imagination to certain other factors. *J. gen. Psychol.*, 1943, 28, 133-141.
6. EIDELBERG, L. Contribution to the study of masturbation fantasy. *Int. J. Psycho-Anal.*, 1945, 26, 127-137.
7. FERENCZI, S. Gulliver fantasies. *Int. J. Psycho-Anal.*, 1928, 9, 283-300.
8. FREUD, S. *Autobiographical study*. London: Hogarth, 1936.
9. FREUD, S. Der Dichter und das Phantasieren. In *Gesammelte Werke*, London: Hogarth, 1941. Vol. VII.
10. FREUD, S. *General introduction to psychoanalysis*. Garden City: Garden City Press, 1920.
11. FREUD, S. Hysterical phantasies and bisexuality. In *Collected papers*. London: Hogarth, 1948. Vol. II.
12. FREUD, S. Hysterische Phantasien und ihre Beziehung zur Bisexualität. In *Gesammelte Werke*. London: Hogarth, 1941. Vol. VII.
13. FREUD, S. Interpretation of dreams. In *The basic writings of Sigmund Freud*. (Ed. by A. A. Brill.) New York: Modern Library, 1938.
14. FREUD, S. *New introductory lectures on psychoanalysis*. New York: Norton, 1933.
15. FREUD, S. Poet and daydreaming. In *Collected papers*. London: Hogarth, 1948. Vol. II.
16. FRIEDMANN, A. Über das Minderwertigkeitsgefühl phantasievoller Kinder. *Z. pädag. Psychol.*, 1932, 33, 273-279.
17. FÜLOP, S. The relationship between fantasy and lying in relation to the body structure of the mentally inferior. *Magyar Psychol. Szemle*, 1929, 2, 58-74. (*Psychol. Abstr.* 1931, 5, 1024.)
18. Green, G. H. *Psychoanalysis in the classroom*. New York: Putnam, 1922.
19. Hart, B. *Psychology of insanity*. New York: Macmillan, 1931.
20. Hesnard, A. Contribution à l'étude des phantasmes érotiques. *Rev. franc. Psychanal.*, 1930-31, 4, 525-548.
21. HULL, C. L. Behavior postulates and corollaries. *Psychol. Rev.*, 1950, 57, 173-180.
22. HULL, C. L. Mind, mechanism, and adaptive behavior. *Psychol. Rev.*, 1937, 44, 1-32.
23. HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
24. HURLOCK, ELIZABETH B., & BURNSTEIN, H. Imaginary playmate: A questionnaire study. *J. genet. Psychol.*, 1932, 41, 380-392.
25. JAEHNER, D. Die Phantasieschwester.

- Z. pädag. Psychol.*, 1930, 31, 371-379.
26. KAMIAT, L. Cosmic phantasy. *Psychoanal. Rev.*, 1928, 15, 210-219.
 27. KOCH, S. Logical character of the motivation concept. *Psychol. Rev.*, 1941, 48, 127-154.
 28. LECKY, P. *Self-consistency: A theory of personality*. New York: Island Press, 1945.
 29. LEHRMAN, P. Significance of fantasy in neurotic behavior. *Med. J. Rec.*, 1927, 126, 342-344.
 30. LÉVY-VALENSI, J. Bovarysme et constitutions mentales. *J. de Psychol.*, 1930, 27, 289-299.
 31. MILLER, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, 38, 89-101.
 32. MOWRER, O. H. Stimulus response analysis of anxiety and its role as a reinforcing agent. *Psychol. Rev.*, 1939, 46, 553-564.
 33. MOWRER, O. H., & KLUCKHOHN, C. Dynamic theory of personality. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Ch. 3.
 34. MOWRER, O. H., & LAMOREAUX, R. R. Fear as an intervening variable in avoidance conditioning. *J. comp. Psychol.*, 1946, 39, 29-50.
 35. MOWRER, O. H., & ULLMAN, A. D. Time as a determinant in integrative learning. *Psychol. Rev.*, 1945, 52, 61-90.
 36. MURRAY, H. A. et al. *Explorations in personality*. New York: Oxford Univ. Press, 1938.
 37. ROGERS, C. Comprehensive theory of personality. Univ. of Chicago, n.d. (Tentative draft.)
 38. ROSENZWEIG, S. Fantasy in psychology. *J. abnorm. soc. Psychol.*, 1942, 37, 40-51.
 39. ROSENZWEIG, S. Systematic intent of clinical psychology. *J. abnorm. soc. Psychol.*, 1949, 44, 3-6.
 40. RUENAUER, H. Wachträume indischer Jugendlicher. *Z. angew. Psychol.*, 1940, 59, 319-350.
 41. SHAFFER, L. F. *The psychology of adjustment*. Boston: Houghton Mifflin, 1936.
 42. SKINNER, B. F. *Behavior of organisms*. New York: Appleton-Century, 1938.
 43. SMITH, THEODATE. Psychology of day-dreams. *Amer. J. Psychol.*, 1904, 15, 465-488.
 44. SPENCE, K. W. Postulates and methods of "behaviorism." *Psychol. Rev.*, 1948, 55, 67-78.
 45. TOLMAN, E. C. *Purposive behavior in animals and men*. New York: Century, 1932.
 46. TSAI, C. Relative strength of sex and hunger motives in the albino rat. *J. comp. Psychol.*, 1925, 5, 407-415.
 47. VAN WATERS, M. "I would rather die than go home!" *Survey*, 1927, 57, 565-569.
 48. VARENDONCK, J. *Psychology of day-dreams*. New York: Macmillan, 1921.
 49. VOSTROVSKY, C. Study of imaginary companions. *Education*, 1894, 15, 383-398.
 50. WARDEN, C. J. *Animal motivation studies*. New York: Columbia Univ. Press, 1931.
 51. WHITE, R. W. Interpretation of imaginative productions. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Ch. 6.
 52. WHITEHEAD, A. N. & RUSSELL, B. *Principia mathematica*. Cambridge: Cambridge Univ. Press, 1925.
 53. WOODWORTH, R. S. *Psychology*. New York: Holt, 1921.

Received November 23, 1950.

THE HISTORY OF THE LEADERLESS GROUP DISCUSSION TECHNIQUE

H. L. ANSBACHER

University of Vermont

The leaderless group discussion is a technique of personality assessment in which a small group are asked to discuss a topic of common interest and are rated individually for various traits by several observers on the basis of discussion behavior. The technique is also known under the names of "group oral performance test" (14), "unsupervised group discussion" (6), and recently "group interview test" (10, 17).

Since its first use in Anglo-Saxon countries during the second World War, the technique has rapidly gained in importance. Bass (3) gives a survey of its spread in England, Australia, and the United States. It is known to have been used for the selection of officer candidates, special military personnel, management trainees and public health officers (3), also of supervisors of special school classes (10), foremen in shipyards (15), and top-level civil servants (21). From Fields (10) we learn that as of 1949, at least ten public and business organizations in the United States employed the technique or experimented with it. Yet its acceptance must have reached further, since Meyer (17) speaks of "the untempered enthusiasm with which the group oral has been accepted in a number of places," and sees the need for warning that "there are no panaceas in the personnel field."

Validity and Reliability of the Technique¹

The technique has gained such wide acceptance so rapidly apparently because it possesses excellent "face" validity and "logical" validity in the sense in which Cronbach (9, pp. 47-55) uses these terms. According to Mandell, "there is overwhelming evidence in regard to the 'face validity' of the group oral as compared with the individual oral." He quotes from three different sources. "The group oral examination method was highly praised . . . The consensus of the participants was that the test was fair to all, interesting.' . . . 'Feelings of confidence in the final choice have on many occasions been expressed by those who have sat on these boards. The reaction of the candidates, also, has been very encouraging.' . . . 'While this technique is un-

¹ The writer is indebted to B. M. Bass, of Louisiana State University, for much of the information in this section.

doubtedly in need of additional testing, we believe that it is superior to the individual interviews used in the past' " (15, p. 183).

As to empirical, correlation validity, few studies are available to date. Mandell (15) found a correlation of .43 between leaderless group discussion (LGD) results and a supervisory judgment test in a study of 84 foremen in two U.S. government shipyards. The correlation with ratings by colleagues and supervisors was .29, which Mandell considers a "substantial validity if one keeps in mind that the reliability of the criterion was probably not higher than .65." Bass and White (4) obtained correlations from .25 to .60 between LGD results and "buddy ratings" of fraternity members. This criterion had a reliability of over .90, and biserial validity correlations were used. Vernon (21) reports an average correlation of .36 between LGD ratings and various measures of performance of top-level civil servants in the British foreign service obtained as much as two years later. The correlations between LGD ratings and a battery of aptitude and achievement tests, heavily loaded with verbal components, averaged around .30. The LGD ratings accounted for over 50 per cent of the variance in the final disposition of the candidates, where assessment and assignment were based on written qualifying examinations, aptitude and achievement test batteries, short talks delivered by candidates, interviews, and pooled committee ratings.

Regarding reliability, only one study has been published which might be mentioned here. Bell and French (5) had 25 students participate in six five-man discussion groups arranged so that each individual met with completely different discussants each session. At the end of each session, the members of a group ranked their four fellow-members in order of preference for discussion leader. The average rank of the five members was then correlated with their average ranks in the other five groups in which they participated. This correlation was found to be .75 on the average.

But research concerning the LGD has only barely begun. Under B. M. Bass, of Louisiana State University, numerous studies on the subject are in progress. Work is also being done at the University of Rochester under L. F. Carter, and at Wayne University under E. T. Raney.

Origin of the Technique

The question of origin arises in connection with a new technique which has within a few years caught the imagination of many. According to accounts given in recent papers (3, 10), it originated in the

British Army during World War II, as part of a much larger program of personnel assessment through group situation tests and other procedures. Regarding group situation tests in general, credit is given to the German Army (3), where, about 1925, such tests were actually used first. But it is not realized that the LGD, specifically, was an integral part of German military personnel selection from the beginning (1, p. 384). The Germans called it *Schlusskolloquium*, because the discussion took place at the end of the entire selection procedure. Sometimes they referred to it as *Rundgespräch*, which means almost literally round-table discussion.

The originator of the technique must be considered to be J. B. Rieffert, who directed German military psychology from 1920 until 1931 (19). Subsequently, he became an industrial psychologist, and never published anything but a few minor papers. He called the LGD simply *Kolloquium*, and sometimes conducted it over the dinner table. Rieffert states: "The round-table discussion as a selection procedure was used in Germany for the first time around 1925 . . . It was introduced by me in connection with the testing of officer candidates of the postwar [World War I] Army; I developed it together with my co-workers . . . Hans Friedländer, later *Dozent* in Berlin, now professor in England; Johannes Rudert, later professor in Leipzig; and Philipp Lersch, later professor in Munich."² According to Rieffert the group discussion was one of a group of four procedures used to obtain samples of problem-solving and social behavior:

1. The command series (*Befehlsreihe*), in which the candidate was asked to carry out complicated orders, shows behavior "toward specific tasks under difficult circumstances."

2. The leadership test (*Führerprobe*), in which the candidate instructed and supervised a group of men in some tasks, shows behavior "toward inferiors."

3. The group discussion (*Rundgespräch*) shows behavior "toward equal partners."

4. The interview (*Exploration*) shows behavior "toward the intellectually superior."

On these behaviors Rieffert based deductions regarding the candidate's attitudes toward work, toward his fellow-men, and toward himself.

Use of the Technique in Germany

Use in the German Armed Forces. Around 1935 German military psychology was split into Army, Navy, and Air Force psychology. The Air Force, according to Fitts (12, p. 155) included a group discussion in the testing of its officer candidates, but we do not know what im-

² Personal communication.

portance was attached to this procedure. From the Army we have the following description of its use by S. Gerathewohl, a former Army psychologist, now in this country.

The round-table discussion was already in use for the selection of officer candidates in 1934 when I became an Army psychologist After the individual interview, the entire group gathered for a final, round-table discussion (in most cases the group actually sitting around a table or in a circle). In Dresden and Breslau where I worked, the discussion was usually started by the psychologist as follows: "Now, that you are finished with all the tests, we shall talk for a while about your experiences and feelings regarding the examination. You can discuss the matter very frankly and express your opinion and criticism without being afraid that this would be to your disadvantage." Thus starting with a topic with which the candidates were very much involved and strengthening the courage of the individual by a social situation, the discussion usually became at once very lively. It was easy to shift from the initial topic of the discussion to such other topics as: officer qualifications, the importance of a strong army, etc. The directive role of the psychologist depended on the spirit of the group, the temperament displayed, and on the general level of the discussion In most cases the discussion was ended with the election of a group leader who—it was assumed—was to represent the group in an affair of honor. This democratic act was performed in the German Army as late as 1939, at least at Assessment Station IV, Dresden, under Lucke.³

It seems that gradually the group discussion was neglected by the Army; a search of the *Wehrpsychologische Mitteilungen*, the house organ of Army psychology, published in monthly issues of 50 to 90 pages from 1939 to 1942, did not yield any article on the technique.

In the Navy the situation was different. According to Mierke, who was in charge of German Naval psychology:

The group discussion gained continuously in importance with us and came to be one of the weightiest procedures It remained a part of the abbreviated officer selection program of the late years of the war and was used occasionally even in the selection of specialists.⁴

Around 1938 Mierke had modified the discussion into an openly controversial one because he had found this more revealing.

The form of defense and attack; the striving for objectivity, for conciliation or for compromise solutions; . . . sense of humor and the lack of it; slyness and rudeness—in short almost all forms of behavior in social relations can become manifest in this procedure. As a rule it will even be necessary to calm heated tempers at the conclusion by asking that the entire affair be regarded only as an amusing intellectual game and that the clash of opinions be not continued elsewhere (18, p. 62).

Controversial topics generally suitable for 18-year-old youths were

³ Personal communication.

⁴ Personal communication.

found to be "For or Against Dancing Lessons" or "Moderate Smoking versus No Smoking." Topics of a military or political nature were avoided on purpose to obtain spontaneous rather than drilled or dishonest behavior.

Civilian use in wartime Germany. With the growing manpower shortage in wartime Germany, psychology experienced a boom in that it was increasingly applied in the nationwide talent searches which were conducted in many fields. The actual selection of talent frequently took the form of selection camps, modeled after the tried military method (2, pp. 610-611). Among the procedures used was the group discussion.

A detailed recommendation of procedure for use in metal-industry camps for the selection of able young workers is to be found in an article by Fischer and Lottmann (11). The camp session, which is to last about ten days, would include three group discussions of about one hour each. The first discussion would refer to occupational goals, the second to some technical problem, the third to impressions gained from a preceding visit to a nearby plant.

Dr. Christel Drey-Fuchs, who worked as a psychologist in trade, commercial, educational, and artistic selection camps, states that one or more group discussions were generally carried out.⁵ There is also a corroborative account by Dr. Helga Schmidt-Oesfeld, a psychologist who had been examined in such a camp in connection with her application for a study grant.⁶

Use of the technique in present-day Germany. Today the group discussion is used in Germany as part of the entrance examination to some teachers colleges. Stückerath (20) conducted an informal survey among teacher-training institutions and found that the group discussion is favored "where for lack of time one is compelled to adhere to the more conventional methods of selection and yet does not want to do without a systematic observation of behavior." Among the topics are: "Are Wars Avoidable?" "Modern Dance," "Alcohol and Nicotine," "Coeducation," and "Corporal Punishment." Factors to be considered in the evaluation are: Is the examinee in rapport with the group? Does he respect the opinion of others? Does he give the other fellow a chance to talk? In all, the group discussion is seen as a test of social attitudes.

Specifically, the group discussion is used at the Teachers College (*Pädagogische Hochschule*) in Kiel under Mierke,⁷ the former director

⁵ Personal communication.

⁶ Personal communication.

⁷ Personal communication.

of Naval psychology, who considers one of the merits of the procedure to be that nonpsychologists can become reasonably well trained in its application. Another teachers college where the group discussion is used in the selection of students is the Pedagogical Institute of Darmstadt at Jugenheim. An American exchange professor, who had a chance to observe these discussions, stated that he was greatly impressed by them and their possibilities and that they were superbly handled by Professor Ruppert. In an interview with the present writer Ruppert explained that the group discussion is used for the appraisal of personality and social attitudes. His preferred topics are "The Ideal Teacher" and "What Difficulties Do I Expect to Encounter in Teaching?" He found it advantageous to seat the candidates in an open circle—rather than around a table—because when deprived of the prop of a table, they are literally and figuratively speaking more open to inspection. In addition to about eight candidates, two to three judges are included in the circle, and these have a table in front of them for note-taking. A judge would occasionally interfere to help a particularly shy individual enter the discussion. Ruppert had been an Army psychologist at one time during the war.⁸

Social-Psychological Implications

We have shown that the group discussion test was introduced in Germany twenty-five years ago and has been holding its place there ever since. While the details of the picture have not been presented heretofore, the basic fact had been mentioned in the American psychological literature at least four times—in 1941 (1, p. 384), in 1942 (16, p. 176), in 1945 (22), and in 1946 (12, p. 155)—in media of widest circulation and often-quoted references, with exception of the 1945 reference, which is only an abstract. The paper by Martin (16, p. 176) devoted half a page to a description of the technique.

How is it to be explained that American psychologists interested in the group discussion have nevertheless remained oblivious to the fact that it was originally used in Germany? One explanation, of course, might be that it was simply overlooked. But might it not also be that the group discussion method, with its aspects of spontaneity, fluidity, and democracy, did least fit into the prevailing simplified conception

⁸ Since this paper was completed, it has become known that the group discussion is being used for the assessment of officer candidates in the newly created West-German emergency police. (Scharmann, T. Bericht über psychologische Untersuchungen bei den Auswahllehrgängen der Bereitschaftspolizei in Traunstein. *Psychol. Rundschau*, 1951, 2, 115.

of the German scene, particularly as it appeared under the Hitler dictatorship? In that case we would have an example in the field of social psychology of the frequently noted mechanism of maintaining an original conception by overlooking elements which would disturb it.

A paper by Jennings (13) gives an eloquent illustration of the mechanism by which stereotypes are protected. She states:

As we study the situation tests used by the military psychologists in Germany under the Nazi regime . . . we note that not one of them allows the individual scope and variety in solutions, nor gives him a chance for personality expression *per se* He is never placed in a setting where he has opportunity to develop a relationship between himself and specific other persons . . . nor in any situations specifically constructed to be meaningful to him as a particular individual Under the Nazi regime, it may very well be that the psychological climate could ill afford to encourage spontaneous expression either in its experimental program or its regime as a whole. (13, p. 191).

Jennings sees complete consistency between psychological methods and political climate.

Actually, however, any social phenomenon is complex and very likely to include inconsistencies. Recently, an entire symposium was devoted to the problem of inconsistency in the social attitudes and the behavior of one and the same individual (8). One of the conclusions reached by Chein was that "The first step toward improving the quality of research in intergroup relations is the awareness of consistency-inconsistency as a pertinent dimension" (7, p. 52). If inconsistency regarding social behaviors is such an important factor on the individual-psychological level, how much more must we be prepared for inconsistencies on the institutional-sociological level of an entire nation. The discovery of an existing inconsistency, while disturbing to stereotypes, is highly welcome to those interested in changing a given attitude or condition, because the inconsistency affords a basis for any attempt to bring about change (7, p. 59).

The discovery of an inconsistency also presents a challenge for closer examination. With regard to the present situation we find that the Nazi regime was not as completely totalitarian as it appeared. Military psychology was not particularly Nazi, and relatively few psychologists were members of Nazi organizations. That is not to say that those in charge were not solid militarists; but the minor psychologists were not necessarily even that. Actually, the type of people military psychology attracted earned it the nickname of "internal emigration." Its partial dissolution during the war may well have been caused by the smoldering conflict with the Nazi party and principles, as many of these psychologists have claimed. In fact, the Air Force, which stood most strongly

under Nazi influence, dismissed its psychologists first; next came the Army. The Navy, which remained most autonomous, retained its psychologists to the end of the war. The Navy was also the most faithful user of the LGD.

Summary

The leaderless group discussion test has received increasing attention in Anglo-Saxon countries since World War II through its very satisfactory face and logical validity. Available correlational research regarding the technique is still scanty, but a considerable amount of such work is in progress. The importance which the technique has achieved has led the author to an outline of its history in Germany from its beginnings in 1925 to the present day. It was used in the German Army and Air Force, but particularly in the German Navy; it was used in Nazi-sponsored selection camps devoted to a widespread talent search; it is used in postwar Germany as part of the entrance examination to certain teachers colleges. The use of such an unstructured, democratic technique does not fit the simplified conception of Germany which still prevails, and this may be one reason why the German priority in the group discussion has heretofore been overlooked.

BIBLIOGRAPHY

1. ANSBACHER, H. L. German military psychology. *Psychol. Bull.*, 1941, 38, 370-392.
2. ANSBACHER, H. L. German industrial psychology in the fifth year of war. *Psychol. Bull.*, 1944, 41, 605-614.
3. BASS, B. M. The leaderless group discussion technique. *Personnel Psychol.*, 1950, 3, 17-32.
4. BASS, B. M., & WHITE, O. L. Validity of leaderless group discussion observers' descriptive and evaluative ratings for the assessment of personality and leadership status. *Amer. Psychologist*, 1950, 5, 311-312. (Abstract.)
5. BELL, B., & FRENCH, R. L. Consistency of individual leadership position in small groups of varying membership. *J. abnorm. soc. Psychol.*, 1950, 45, 764-767.
6. BRODY, W. Judging candidates by observing them in unsupervised group discussion. *Personnel J.*, 1947, 26, 170-173.
7. CHEIN, I. The problem of inconsistency: A restatement. *J. soc. Issues*, 1949, 5, No. 3, 52-61.
8. CHEIN, I., et al. (Eds.) Consistency and inconsistency in intergroup relations. *J. soc. Issues*, 1949, 5, No. 3.
9. CRONBACH, L. J. *Essentials of psychological testing*. New York: Harper, 1949.
10. FIELDS, H. The group interview test: Its strength. *Publ. Personnel Rev.*, 1950, 11, 139-146.
11. FISCHER, G. H., & LOTTMANN, W. Menschenbild, Auslese und Erziehung. Beitrag VIII: Zur Methodik im beruflichen Ausleselager. *Z. pädag. Psychol.*, 1943, 44, No. 5, 85-93.

12. FITTS, P. M. German applied psychology during World War II. *Amer. Psychologist*, 1946, 1, 151-161.
13. JENNINGS, HELEN H. Military use of sociometric and situation tests in Great Britain, France, Germany and the United States. *Sociometry*, 1949, 12, 191-201.
14. MANDELL, M. M. The group oral performance test. *Publ. Personnel Rev.*, 1946, 7, 209-212.
15. MANDELL, M. Validation of group oral performance test. *Personnel Psychol.*, 1950, 3, 179-185.
16. MARTIN, A. H. Examination of applicants for commissioned rank. In *L. Farago (Ed.), German psychological warfare* (1st definitive Ed.) New York: Putnam, 1942. Pp. 171-178.
17. MEYER, C. A. The group interview test: Its weakness. *Publ. Personnel Rev.*, 1950, 11, 147-154.
18. MIERKE, K. Psychologische Diagnostik. In N. K. Ach (Ed.), *Lehrbuch der Psychologie: Praktische Psychologie*. Bamberg: C. C. Buchner, 1944. Vol. 3, pp. 1-79.
19. SIMONEIT, M. Vom Werden der deutschen Wehrmachtpsychologie; ein geschichtlicher Rückblick. *Wehrpsychol. Mitt.*, 1940, 2, Suppl. No. 2.
20. STÜCKRATH, F. Der gegenwärtige Stand der Ausleseverfahren für den Lehrerberuf. *Bildung und Erziehung*, 1949, 2, 501-516.
21. VERNON, P. E. The validation of Civil Service Selection Board procedures. *Occup. Psychol., Lond.* 1950, 24, 75-95.
22. WINZER, H. A. Abstract of Fischer, G. H., & Lottmann, W. Menschenbild, Auslese und Erziehung. Beitrag VIII. Zur Methodik im beruflichen Ausleselager. *Z. pädag. Psychol.*, 1943, 44, No. 5, 85-93. (*Psychol. Abstr.*, 1945, 19, No. 3465.)

Received January 20, 1951.

COMPUTATION OF THE LEVEL OF SIGNIFICANCE IN THE F -TEST

C. J. BURKE

Indiana University

Frequently, the psychological experimenter expresses dissatisfaction with the ordinary table for interpreting F -ratios. This table gives the values of the F -ratio which are required for significance at the .01 and .05 levels, but, unlike the t and χ^2 tables, interpolations for values below the .05 level are not possible. Faced with a value of F which is not significant at the .05 level, the experimenter often wishes to estimate its actual significance. If it is significant at the 8 per cent level, for example, further experimentation may be indicated, but if it is barely significant at the 30 per cent level, the experimenter might discontinue working with the particular variables involved.

One way of resolving this state of affairs would be the provision of an adequate F -table. The F -table is essentially three-dimensional and cannot be reduced to two dimensions—its adequate representation would require a book, rather than a page. In spite of the extensive labors that would be required, it might be of use to tabulate the entire F -distribution for such numbers of the degrees of freedom as occur rather frequently, were it not for the fact that an extensive table of an intimately related distribution already exists. Pearson (2) has tabulated the beta-distribution, and the book in which this distribution is represented is rather widely available in university libraries. A simple computational procedure enables one to assign a level of significance to any value of F by means of the tabulated beta-distribution.

The intimate relation between these two distributions has long been known to professional statisticians, and the underlying theory relating them is presented in rather widely used treatises by Cramér (1), Wilks (3), and many others. In the present paper, the results of this theory will be summarized in a form which is mathematically less formidable than the usual presentations. Following the presentation of the theory, the detailed steps in the calculations are illustrated by means of several examples. The psychological statistician who is faced with the problem of calculating the level of significance of a value of F and who is not interested in the underlying theory can use the examples as a model for his calculations without reading the theoretical section.

THEORY

With the quantity which we compute as F is associated a probability distribution from which the probability that a value of F selected at random lies between any two given values can be calculated. We shall symbolize this distribution in its cumulative form by $G_{r_1, r_2}(F)$, where r_1 and r_2 are the number of degrees of freedom respectively associated with the variances in the numerator and denominator of the F -ratio. (this means that r_1 will actually occur in the denominator and r_2 in the numerator of the F -ratio, since that ratio is obtained by dividing variance estimates). If we are given a specific value of F , denoted by F_0 , with specified values of r_1 and r_2 , the probability that any value of F selected at random will be smaller than F_0 can be calculated from:

$$P(F \leq F_0) = G_{r_1, r_2}(F_0). \quad [1]$$

The ordinary F -table gives the values of F_0 for each combination of r_1 and r_2 values which makes the probability equal to .99 or .95. If we wish to work at levels of confidence other than .01 or .05, the table is of little use.

Because the basic definition of F involves the ratio between two values of chi-square, the reciprocal of F also has the F distribution, but with the order of r_1 and r_2 reversed. Hence,

$$P\left(\frac{1}{F} \leq \frac{1}{F_0}\right) = G_{r_2, r_1}\left(\frac{1}{F_0}\right). \quad [2]$$

But, from the nature of reciprocals,

$$P\left(\frac{1}{F} \leq \frac{1}{F_0}\right) = P(F \geq F_0) = G_{r_2, r_1}\left(\frac{1}{F_0}\right). \quad [3]$$

Also, since the distribution is continuous,

$$P(F \geq F_0) = P(F > F_0) = 1 - P(F \leq F_0). \quad [4]$$

From [1] and [4],

$$P(F \geq F_0) = 1 - G_{r_1, r_2}(F_0). \quad [5]$$

The combination of [3] and [5] yields a relation between the two cumulative distribution functions:

$$G_{r_2, r_1}\left(\frac{1}{F_0}\right) = 1 - G_{r_1, r_2}(F_0). \quad [6]$$

Subsequent use will be made of equation [6]. It is well known (3,

pp. 114-115) that there exists an intimate connection between the F -distribution and the incomplete beta-function tabulated by Pearson. If we sample at random from an F -distribution and convert each value of F obtained to a value of x according to the formula

$$x = \frac{\frac{r_1}{r_2} F}{1 + \frac{r_1}{r_2} F}, \quad [7]$$

the values of x obtained will range between 0 and 1 and will have the distribution tabulated by Pearson. When a given value F_0 has been selected, a value of x , denoted by x_0 , can be computed from

$$x_0 = \frac{\frac{r_1}{r_2} F_0}{1 + \frac{r_1}{r_2} F_0}. \quad [7']$$

Then

$$P(x \leq x_0) = P(F \leq F_0). \quad [8]$$

The probability on the left of [8] can be obtained from Pearson's tables. The cumulative distribution function of x is written symbolically as $I_{p,q}(x)$ ¹ so that

$$P(x \leq x_0) = I_{p,q}(x_0) \quad [9]$$

where

$$p = \frac{r_1}{2} \quad \text{and} \quad q = \frac{r_2}{2}. \quad [10]$$

From [1], [8], and [9],

$$G_{r_1, r_2}(F_0) = I_{p, q}(x_0). \quad [11]$$

Pearson's tables give values of $I_{p,q}(x)$ for various values of p , q , and x , such that $q \leq p$. This restriction on values of p and q means that we must consider two cases.

Case 1. $r_1 \geq r_2$, ($p \geq q$). We are given a value F_0 and seek to compute

¹ The notation here employed differs slightly from that used by Pearson, but the correspondence between the two notational systems is obvious and should cause no confusion.

its level of significance $[1 - G_{r_1, r_2}(F_0)]$. We calculate p , q , and x_0 from [10] and [7']. Then we look up the value of $I_{p, q}(x_0)$ in Pearson's tables. This value is $G_{r_1, r_2}(F_0)$ according to [11] and hence we obtain the desired level of confidence by subtracting this value from unity.

Case 2. $r_1 < r_2$, ($p < q$). Again, we are given a value of F_0 and desire the value of $1 - G_{r_1, r_2}(F_0)$. If we follow the method outlined in case 1, we are unable to find $I_{p, q}(x_0)$ in Pearson's table since values for $p < q$ are not tabulated. To get around this difficulty, we make use of the fact that the reciprocal of F also has the F distribution. We calculate x_0 according to:

$$x_0 = \frac{\frac{r_2}{r_1} \frac{1}{F_0}}{1 + \frac{r_2}{r_1} \frac{1}{F_0}}. \quad [12]$$

For the reader who is familiar with the beta-function, the significance of the change from formula [7'] to formula [12] will be clarified by noting that, if x is calculated from formula [7'], the quantity on the right of [12] gives $1 - x$. Thus, changing r_1 and r_2 in the F -distribution changes p and q in the beta-distribution.

Then

$$G_{r_2, r_1}\left(\frac{1}{F_0}\right) = I_{p, q}(x_0) \quad [13]$$

where

$$p = \frac{r_2}{2} \quad \text{and} \quad q = \frac{r_1}{2}. \quad [14]$$

With equation [14], $p \geq q$, and the value of $I_{p, q}(x)$ can be found in Pearson's table. From [13] and [6], we have

$$1 - G_{r_1, r_2}(F_0) = I_{p, q}(x_0) \quad [15]$$

so that the value of $I_{p, q}(x_0)$ is taken directly as the level of significance.

COMPUTATIONAL EXAMPLES

Case 1. $r_1 > r_2$. *Example 1.* We are given a value of $F_0 = 1.30$ with degrees of freedom $r_1 = 30$ and $r_2 = 19$ associated with the numerator and denominator, respectively. According to [7'], we calculate x_0 as

$$x_0 = \frac{\frac{30}{19}(1.30)}{1 + \frac{30}{19}(1.30)} = .673,$$

and $p=15$, $q=9.5$ from [10]. From Pearson (2, p. 240), we obtain

$$I_{15,9.5}(.670) = .713$$

$$I_{15,9.5}(.680) = .748.$$

By interpolation

$$I_{15,9.5}(.673) = .724.$$

Hence the level of significance is $1 - .724 = .276$.

Example 2. Given $F_0=1.50$, $r_1=27$, $r_2=25$, find the level of significance.

$$x_0 = \frac{\frac{27}{25}(1.50)}{1 + \frac{27}{25}(1.50)} = .618$$

from [7'], and $p=13.5$, $q=12.5$ from [10]. From Pearson, we obtain on page 278 ($q=12$)

$$I_{13,12}(.610) = .815$$

$$I_{14,12}(.610) = .765$$

$$I_{13,12}(.620) = .842$$

$$I_{14,12}(.620) = .796$$

and on page 288 ($q=13$)

$$I_{13,13}(.610) = .870$$

$$I_{14,13}(.610) = .829$$

$$I_{13,13}(.620) = .891$$

$$I_{14,13}(.620) = .855.$$

At this juncture, interpolation with respect to three variables is necessary. To obtain really accurate results, higher-order interpolation formulas should be used, but for most purposes, linear interpolation will yield sufficient accuracy. Interpolating first with respect to q , we obtain

$$I_{13,12.5}(.610) = .842$$

$$I_{14,12.5}(.610) = .797$$

$$I_{13,12.5}(.620) = .866$$

$$I_{14,12.5}(.620) = .826.$$

Interpolation with respect to p yields

$$I_{12.5,12.5}(.610) = .820$$

$$I_{12.5,12.5}(.620) = .846.$$

It should be noted that the interpolations with respect to p and q can be done in one step:

$$I_{12.5,12.5}(.610) = \frac{.815 + .870 + .765 + .829}{4} = \frac{3.279}{4} = .820$$

$$I_{12.5,12.5}(.620) = \frac{.842 + .796 + .891 + .855}{4} = \frac{3.384}{4} = .846.$$

Finally, we interpolate with respect to x_0 to obtain

$$I_{12.5,12.5}(.618) = .841.$$

Hence, the level of significance is $1 - .841 = .159$.

Case 2. $r_1 < r_2$. *Example 3.* We are given a value of $F_0=1.50$ with $r_1=25$ and $r_2=75$. Using equation [12], we have

$$x_0 = \frac{\frac{75}{25} \left(\frac{1}{1.50} \right)}{1 + \frac{75}{25} \left(\frac{1}{1.50} \right)} = .667$$

and from [14], $p=37.5$, $q=12.5$. From Pearson's tables on page 285 ($q=12$), we obtain

$$I_{37,12}(.660) = .0678$$

$$I_{38,12}(.660) = .0564$$

$$I_{37,12}(.670) = .0886$$

$$I_{38,12}(.670) = .0748$$

and on page 294 ($q=13$), we obtain

$$I_{37,13}(.660) = .1027$$

$$I_{38,13}(.660) = .0870$$

$$I_{37,13}(.670) = .1313$$

$$I_{38,13}(.670) = .1126.$$

Interpolation with respect to p and q yields:

$$I_{37.5,12.5}(.660) = \frac{.0678 + .0564 + .1027 + .0870}{4} = .0785$$

$$I_{37.5,12.5}(.670) = \frac{.0886 + .0748 + .1313 + .1126}{4} = .1018.$$

Finally, interpolation with respect to x_0 gives

$$I_{37.5,12.5}(.667) = .0948.$$

In case 2, no subtraction is necessary as this figure gives the level of significance directly.

SUMMARY

In this paper, the well-known theory by means of which the significance of any obtained F -ratio can be obtained from tables of the incomplete beta-function is summarized. Computational examples are presented which can serve as models for investigators interested in determining the exact significance of values which do not occur in the ordinary table for F .

BIBLIOGRAPHY

1. CRAMÉR, H. *Mathematical methods of statistics*. Princeton: Princeton Univ. Press, 1946.
2. PEARSON, K. *Tables of the incomplete beta-function*. London: Biometrika office, 1934.
3. WILKS, S. S. *Mathematical statistics*. Princeton: Princeton Univ. Press, 1944.

Received November 22, 1950.

ON THE USE OF LATIN SQUARES IN PSYCHOLOGY

QUINN McNEMAR

Stanford University

The purpose of this note is to attempt a clarification of two aspects of the latin square design which seem inadequately treated in the expositions written primarily (2, 4, 5, 11) or partly (6, 10) for psychologists. First we shall consider briefly the types of situations where the latin square design might be useful, and second we shall point out the statistical assumptions which must be met. Then we shall face the implications of the latter for the former.

To facilitate the discussion, suppose the agricultural situation represented by the accompanying square, in which the letters stand for

		<i>Rows</i>			
		1	2	3	4
<i>Columns</i>	I	A	D	B	C
	II	B	A	C	D
	III	C	B	D	A
	IV	D	C	A	B

four different treatments so arranged in a field plot that each treatment occurs once in each row and once in each column. The object is to average out possible fertility differentials from row to row and from column to column. With soil heterogeneity thus balanced, the experiment is obviously under better control, hence the results should have greater precision; this greater precision is reflected in an error term which is the residual *after* variations due to row differences and to column differences (also treatment effects) have been deducted.

Since the design need not be earth-bound, it is only natural that psychologists should adopt it for those situations where they wish to balance out the effects of sources of variation not experimentally controllable. The letters are assigned to "treatments" (dosages, methods, conditions, etc.) while the rows and columns typically stand for subjects and order of testing. Variations due to subjects and due to order (practice or fatigue) are consequently balanced as regards the treatments, and the statistical analysis capitalizes on this fact in that the error term used for the *F* ratio does not include variation associated with subjects and with order.

The latin square was not invented for the purpose of providing a means for taking care of variation due to easily controlled variables,

e.g., illumination. However, the design can be used for situations in which the columns and rows need not represent uncontrollable sources of variation. Either the rows or the columns or both can stand for factors which are deliberately varied for the sake of testing hypotheses regarding their effects. That the latin square design may be used instead of the so-called factorial design seems to have been overlooked in the expositions for psychologists. As early as 1937, Fisher, in his *The Design of Experiments* (3, section 35.1), mentioned the possibility. More recently the idea has appeared in books by Kendall (7), Cochran and Cox (1), and Mood (9). Stated very briefly, a latin square design with 16 properly arranged observations may be used in place of a complete three-way analysis of variance (factorial) design requiring 64 observations; 25 observations may be used instead of the 125 needed in a three-way analysis with five levels for each classification (or factor); and so on.

Obviously, each factor must involve the same number of levels to permit the substitution of a latin square design for a complete factorial design, and equally obvious is the fact that letters in such squares stand for levels on one of the factors. This use of a latin square allows the testing of hypotheses regarding all three factors. As usual, the error term is free of variation due to the three factors under study. Unlike the complete three-way analysis of variance, the interactions cannot be tested. The chief advantage of the use of a latin square in place of a factorial design is that fewer observations are required—an important consideration when the securing of an observation is costly.

It has been implied above that the latin square is permissible for a mixed situation—either the rows (or columns) standing for a factor to be investigated while the columns (or rows) stand for an uncontrollable variable the effect of which needs to be balanced. Accordingly, the latin square would seem to be a flexible design, which it would be if it were not for the matter of assumptions.

Aside from the usual assumption that the observations are from normally distributed populations with equal variances, *it is also assumed that all interactions are zero*. For some reason or other the advocates of the use of the latin square in psychological research have passed over this second and fundamental assumption. Though Grant (5) comes nearest to pointing out this assumption when he says that the interactions are confounded, cannot be assessed, and “may influence the size of the error term,” he fails to make it explicit. One searches in vain for mention of this assumption in the 30-page chapter on latin squares in Edwards’ recent book (2). The assumption is hinted at in Fisher’s

Design (3, section 35.1), while it is directly implied in the mathematical formulations of Cochran and Cox (1, pp. 41-42), of Mann (8, p. 77), of Mood (9, p. 340), and of Wilks (12, p. 190). Mood also puts it in near Basic English: "all interactions are assumed to be zero."

What is the possible consequence of failure to meet this assumption? From the viewpoint of statistical theory it simply means that F 's from latin squares do not follow the F distribution—too many "significant" F 's will be obtained when the assumption is not met. This will happen because the residual term involves an admixture of ordinary error and any two-way interaction that is present, which interaction will of course be larger than the ordinary error component, but the combination of the two sources will tend to yield a residual which is smaller than the interaction that properly should be used as the denominator for F .

What of the likelihood that the assumption will not hold for psychological variables? First, consider the possible use of a latin square in lieu of a complete factorial design. One need only glance through the *Journal of Experimental Psychology* to learn that in about half the studies involving the testing of interactions, significant interactions between the factors emerge. Second, consider the commonly used situation in which either the rows or the columns stand for individuals. Since there is nearly always an interaction between individuals and factors, it follows that the assumption will nearly always be violated in such situations.

Faced with these facts, we are forced to the inescapable conclusion that the latin square design is seldom appropriate in psychological research. *It is defensible only in those rare instances when one has sound a priori reasons for believing that the interactions are zero.*

BIBLIOGRAPHY

1. COCHRAN, W. G., & COX, G. M. *Experimental designs*. New York: Wiley, 1950.
2. EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
3. FISHER, R. A. *The design of experiments*. (2nd Ed.) Edinburgh: Oliver and Boyd, 1937.
4. GARRETT, H. E. & ZUBIN, J. The analysis of variance in psychological research. *Psychol. Bull.*, 1943, 40, 233-267.
5. GRANT, D. A. The latin square principle in the design and analysis of psychological experiments. *Psychol. Bull.*, 1948, 45, 427-442.
6. JOHNSON, P. O. *Statistical methods in research*. New York: Prentice-Hall, 1949.
7. KENDALL, M. G. *The advanced theory of statistics*. London: Griffin, 1948. Vol. II.
8. MANN, H. B. *Analysis and design of experiments*. New York: Dover, 1949.

9. MOOD, A. M. *Introduction to the theory of statistics*. New York: McGraw-Hill, 1950.
10. PETERS, C. C., & VAN VOORHIS, W. R. *Statistical procedures and their mathematical bases*. New York: McGraw-Hill, 1940.
11. THOMSON, G. H. The use of the latin square in designing educational experiments. *Brit. J. educ. Psychol.*, 1941, 11, 135-137.
12. WILKS, S. S. *Mathematical statistics*. Princeton: Princeton Univ. Press, 1943.

Received November 29, 1950.

THE GENETICS OF SCHIZOPHRENIA: REPLY TO PASTORE¹

LEWIS A. HURST

Weskoppies Hospital, Pretoria, South Africa

The two main technical sources of Pastore's errors in his criticisms of Kallmann's *The Genetics of Schizophrenia* (4) are:

1. Ignorance of the more recent experimental developments in genetics and the conceptual framework of physiological genetics erected thereon.

2. Ignorance of certain special statistical techniques relevant to genetic and population studies evolved chiefly in Germany.

The acclaim accorded to Kallmann's work (1, 5) by Hogben and Haldane at the International Congress of Genetics at Edinburgh in 1939, in the face of their formerly contrary preconceptions, should give one pause before accepting Pastore's picture of Kallmann as a simple-minded medico barging clumsily and unwarily into the domains of science and statistics. Hogben and Haldane are people who *do* know about modern physiological genetics and the statistical methods evolved in connection with genetic and population problems.

Pastore's attitude gains a spurious impetus from the tendency current in "advanced" circles to identify gratuitously a belief in human heredity with a reactionary attitude and that in the all-powerful influence of environment with a progressive attitude. Elsewhere (2, 3), I have sought to indicate the scientifically unacceptable factors of a personal and professional nature entering into this type of environmentalism, as well as the illegitimate extension, by analogy, of Freudian concepts from hysteria to schizophrenia, and the guilelessness of the implied claim of Watson, advanced without any experimental verification, as to the possibility of making a mental defective or genius of an identical individual by implanting different sets of conditioned reflexes.

Before proceeding to a detailed reply to Pastore's criticisms of Kallmann, I feel that there is an onus on me to elaborate and clarify my allegations against him on the score of ignorance of genetics and of certain special statistical techniques evolved in relation to genetics and population studies.

The history of genetics may be divided into three stages:

1. The first is that associated with the name and work of Mendel, in which

¹ PASTORE, N. The genetics of schizophrenia: A special review. *Psychol. Bull.*, 1949, 45, 285-302.

the genes or hereditary units were merely hypothetical entities (whose anatomical basis was unknown) postulated to explain the ratios obtained in cross-breeding experiments.

2. The second stage, associated with the name and work of T. H. Morgan, has as its salient feature the location of Mendel's hypothetical units on a physical substrate, the chromosomes, and in certain species the construction of maps in which the genes responsible for specified traits were assigned positions relative to one another on the individual chromosome pairs.

3. The third stage, that of modern physiological genetics, is associated with the name and work of Goldschmidt. Goldschmidt's experimental developmental studies drew attention to the following fact: the biochemical processes originating from a particular gene pair or set occur in the environment of the rest of the organism, which is the result of the biochemical processes originating from all the other genes of that organism. It is clear that as a result of chemical interaction the processes initiated by the gene pair under consideration may be furthered, inhibited, or modified by the processes deriving from the other genes. Hence the recognition that the genotype, i.e., the sum total of all the component genes of the fertilized ovum, is seldom fully manifested in the phenotype, defined by Kallmann as "manifest features of an organism, representing the end product of the development and appearance of all inherited characters in the individual." Hence, also, the concept of modifying genes or modifiers; and the concept, so very important in modern genetics, of *penetrancy*, which denotes the percentage of cases in which the trait associated with a particular pair or set of genes becomes manifest in the developed organism. The obverse of this concept of penetrancy is the modern tendency in genetics to speak of genetic predisposition, which carries with it a recognition that, few traits being fully or 100 per cent penetrant, the possession by an organism of the requisite complement of genes for a given trait does not imply the invariable appearance of the trait in the fully developed organism, but appearance only in a certain percentage of cases. In the sphere of the clinical entity with which we are concerned, viz., schizophrenia, the concept of constitution (and constitutional resistance) has, as its genetic correlate, modifying genes, which reduce the penetrancy of the single recessive gene pair responsible for the hereditary predisposition to schizophrenia from 100 per cent to about 70 per cent, as Kallmann has shown.

It is clear from Pastore's article that he is living mentally in stage 1 of the history of genetics, not in stage 3, modern physiological genetics, and that is why he fails to understand Kallmann's work. This appears, *inter alia*, from his naive reference, of which he is oblivious of any need for qualification, to "Kallmann's Mendelian outlook."

As regards the statistical methods evolved in relation to human genetics, a full exposition and application to schizophrenic material is to be found in Kallmann's *The Genetics of Schizophrenia*. They are concerned with expectancy figures as opposed to net figures. Here we can only name the methods and indicate their general rationale. All the methods take cognizance of the fact of whether the relative in

question has not reached, has reached, or has passed the danger period for the manifestation of the trait under consideration. In the case of schizophrenia this is taken as from the fifteenth to the forty-fourth year. In exposition of Weinberg's abridged method Kallmann says: "Accordingly, we counted only half of persons who were between the ages of fifteen and forty-four at the end of our statistical control. All persons who were forty-five or older were counted in full, while those who dropped out of our statistics before the age of fifteen were omitted entirely." The Ilse method and morbidity statistics claim greater accuracy than Weinberg's abridged method in that they do not postulate a uniform age distribution within the danger period of manifestation, but there are serious disadvantages in Ilse's method which lead Kallmann after careful analysis to rank it as inferior to Weinberg's abridged method. The detailed mathematical exegesis given by Kallmann in connection with these methods (especially p. 139²), and also Weinberg's double proband method, and Schulz's double case method based on Bernstein's *a priori* method (especially p. 147), accords ill with the simple-minded medico theory. In Weinberg's proband method the principle is adopted that "a proband can be included in the estimate only if he also appears in the survey as the sibling of a proband, while, conversely, the sisters and brothers of a proband should not be selected as probands merely because they are carriers of the schizophrenic trait." Weinberg's double proband method "employs only series of siblings with at least two probands, counting them by means of the usual proband method, as a double proband entity and reckoning their other siblings singly."

Similarly, the double case method is based on Bernstein's *a priori* method, which is used for the separate estimate of the frequency of the given trait both in the series of siblings with at least two trait carriers, and in the total sibling material. The principle of this method rests on the assumption that the hereditary quality of a supposedly transmitted trait, in general occurring rarely, may be viewed as demonstrated if it is found in several individuals of a series of siblings; while there is a stronger possibility of exogenous origin if it is manifested in only one child of a series. Accordingly, if the percentages for the frequency of an hereditary trait in the total survey agree with the figures which are obtained by the same statistical method for the series of siblings with two or more trait carriers, it is highly probable that the proband material under investigation is biologically homogeneous and the given trait is conditioned by heredity.

² The page references cited in this article are to Kallmann's, *The Genetics of Schizophrenia* (4).

Kallmann outlines the mathematical techniques evolved by Weinberg and by Schulz to deal with the factors complicating the *a priori* method in the case of schizophrenia; these factors are that "we are dealing with a late developing and not completely penetrant trait, and a material which comprises a collection of diseased individuals and is not selected directly from the general population." In the section "The Genetic Relation between Schizophrenia and Tuberculosis" (chap. VII), the *G* and *T* methods are differentiated. In the former the tuberculosis mortality rates are related to the totals of observed individuals, while the *T* method deals only with the sum of all deceased persons as a corrected rate of reference.

This brief survey will give some notion of the diversity and complexity of the statistical methods employed by Kallmann, of which his critic shows little if any grasp.

It is now time to answer the detailed criticisms preferred against Kallmann's work.

Establishing the Diagnosis

1. The necessary limitations of a study in human heredity designed to include adequate numbers of ancestors and descendants is appreciated as clearly by Kallmann as by his critic. Kallmann, moreover, emphasizes the need for further studies in many directions.

2. The clinical records of a leading German metropolitan mental hospital at both extremes of the quarter-century referred to were of a standard to permit satisfactory objective assessment, by a modern investigator, of the diagnoses.

3. The clinical content of schizophrenia was identical at the beginning and end of the quarter-century referred to; the only change was that of name—from dementia praecox to schizophrenia.

4. Institutionalization is by no means a necessary feature in the diagnosis of schizophrenia.

5. The exact clinical status and nature of the "doubtfuls" is clearly portrayed for those who take the trouble to read Kallmann's book (especially p. 15). Kallmann's scientific honesty is so strict that his "doubtfuls" probably contain a high proportion of schizophrenics.

6. No statistical error is introduced by Kallmann's division into "definites" and "doubtfuls"; the distinction is carefully maintained both in the tables and in the text, for the first-hand inspection of the readers, and the "doubtfuls" were not included in the calculation of final figures.

7. Whether the "doubtfuls" will become "definites" is an idle question, indicating a lack of understanding of the statistical methods used and of scientific procedure in general. To quote from Kallmann (p. 15): "However, this circumstance did not lead to their being reinstated in any of the principal categories after they had once been dropped for clinical reasons."

8. As regards the statistical points introduced irrelevantly into this section

by Pastore, the criticisms that eight of the secondaries have been classified as probands and that three secondaries were counted twice, are refuted by an understanding of the statistical techniques so lucidly set forth in Kallmann's book.

9. From my own prolonged observation of Kallmann's methods in the course of trips with him in New York State, I can vouch for the fact that his investigation of spouses and other relatives of probands is as thorough, and that his diagnostic procedure is as rigorous, as in the case of probands themselves—resulting in complete diagnostic uniformity.

Sampling

1. Perusal of *The Genetics of Schizophrenia* and a study of the tables will convince the reader that Kallmann includes in his sample the relevant sample characteristics, such as the distribution according to the form of schizophrenia, which his critic alleges he omits.

2. The belief that aged schizophrenics who have been institutionalized for decades have remissions shows Pastore's ignorance of the clinical aspect of the entity under consideration. Any effect that such a factor could have is thus precisely nil. Even if such a factor existed, the reason for Pastore's regarding it as a selective factor is obscure.

3. Kallmann excluded only cases in which exogenous factors played a major part, where the factor was either of a physical nature (e.g., alcohol) and the psychosis was clearly best classified in terms of the physical etiological agent (e.g., alcoholic psychosis), or where he had to do with a clear psychoneurosis.

4. Kallmann's scientific honesty in dropping forty cases because of "doubts regarding the unassailable certainty of the original diagnosis of schizophrenia" is again turned against him by his critic. An answer to the criticism appears in the text of Kallmann's book: "The investigations of their heredity and fertility followed the same lines as in the other proband groups, and even produced series of secondary cases" (p. 15).

5. That children can only originate from fertile parents is an axiom, not a selective factor. A thorough investigation into illegitimate fertility was made, and a comparison with absolute and legitimate fertility was drawn (pp. 51-66). To have sought to study "taint" in ancestry and "taint" in siblings of unmarried probands would have thrown no additional light on the nature-nurture problem in schizophrenia, and the conclusions for illegitimate children would obviously have been less reliable—smaller numbers and less certainty of parentage—than for the legitimate children.

6. The reasons for the high death rate of the children of probands below the age of 20 (chiefly due to high death rate from tuberculosis—five times that of the general population in the second decade of life) are considered by Kallmann, and the statistical consequences carefully evaluated and allowed for.

7. Two thousand one hundred and twenty proband children, including 111 secondary cases, are surely a sufficient number for statistical purposes, as a fourfold subdivision (into the four standard subgroups of schizophrenia) is the maximal subdivision undertaken by Kallmann in his statistical analysis. On page 21 Kallmann explicitly recognizes the inadvisability of dividing the material into too many subgroups, to avoid numbers too low for statistical reliability.

Tabular and Statistical Presentation of Data³

1. Enough has already been said to indicate how well versed Kallmann is in statistical method, notably the specialized statistical techniques applicable to human genetic problems.

2. Although I have studied Kallmann's tables (including Table 10) with extreme thoroughness, I am unable to agree that there are any incorrect titles or percentages with a wrong base. (1)

3. Pastore's drawing attention to the discrepant distributions of the number of schizophrenic children, according to the form of schizophrenia in the proband, derived from the data in Tables 34 to 37 and those in chapter V, rests on his having fallen into the grave error, which he perpetrates so frequently, of failing to discriminate between net and expectancy figures.

4. The numbers included in the subgroups are invariably given, in addition to percentages.

5. Where, in certain tables, cell entries are small, they are never used for establishing statistically significant similarities or differences.

6. Most careful study has failed to reveal to me any instances of the alleged "overlapping categories." They are a fallacious postulate of Pastore, as will be explained later.

7. Two cases, and not one as alleged by Pastore, are characterized as incipient by Kallmann (case 60, p. 182, and case 89, p. 191). Perusal of the clinical descriptions of these two cases provided by Kallmann indicates clearly that although not advanced cases they are certainly definite schizophrenics.

8. The criticism that in selecting final expectancy figures, where differential choice arose, Kallmann consistently chose the higher figure and that he rejected the more accurate method on the grounds that it was more complex is simply untrue. Kallmann's criticism of Ilse's method is on purely logical and mathematical grounds, which he sets out, including Strömngren's arguments for finding Ilse's technique unsatisfactory (p. 142).

9. Kallmann sets forth clearly the expectancy figures for the three main statistical techniques employed, both in the body of chapter IV and in his conclusions (conclusion 6, p. 163). Similarly, the reader is given, at first hand, figures regarding "definites" and "doubtfuls" on which to base an opinion. For his own decision on these points Kallmann gives clear logical and mathematical reasons.

10. Pastore's claim from his own discrepant distribution as to the subgroups of schizophrenia from that given by Kallmann, and his allegation that there is a discrepancy between the biographical and tabular information on the incidence of children of female simple schizophrenics, are based on his failure to distinguish between net and expectancy figures. Pastore generalizes from the alleged error in the expectancy table for children of female simple schizophrenics, to the unreliability of other of Kallmann's tables, when the error is really his!

Evaluation of Kallmann's Major Conclusions⁴

Although certain of Kallmann's major conclusions are included in his critic's list, the principle of choice appears to have been that they

³ The numbering in this section does not correspond to Pastore's.

⁴ The numbering in this section corresponds to Pastore's, with an additional point 7 in reply to his peroration.

are those the critic thought he had refuted, rather than those resulting from systematic selection of the major conclusions of the work.

1. Kallmann makes it quite clear, in the instance alluded to, that the 10.4 figure refers to the number of probands who had a schizophrenic parent and not to the incidence of schizophrenia among the ancestors. On p. 164, in conclusion 14, he writes: "No more than 10.4 per cent of the proband-parents could be ascertained to have been definitely schizophrenic."

2. In the matter of schizophrenic children in the *S* parental subgroup, Pastore has once again erred through confounding expectancy with net figures.

3. Nuclear and peripheral groups of schizophrenia. (a) In view of Pastore's error in 2, his criticism, based on this error, that the distinction between the nuclear and peripheral categories becomes confused, falls away. (b) Once again in this section Pastore confuses net and expectancy figures. (c) The attempt to explain the lower numbers of proband children in the peripheral as compared with the nuclear group in the environmentalistic terms of the lesser impairment of "family structure" resulting from the later breakdown of the schizophrenic parent, is merely a conjecture without any factual verification. In contrast to this, Kallmann's biological interpretation is based on detailed empirical evidence. Moreover, it finds confirmation by interlocking with other exact findings in the fields of genetics and constitution in schizophrenia. A further point is that Kallmann's study as a whole proves that environmentalistic factors of the type here mentioned cannot cause schizophrenia, either of the nuclear or of the peripheral variety.

4. The statistical and diagnostic points referred to by Pastore in treating of final expectancy figures in various groups of blood relationship have already been dealt with. A further misstatement introduced by Pastore is that the figure of 16.4 per cent for proband children is inflated by the inclusion of "doubtful" cases. That this is not so can be confirmed by the reader for himself by studying Table 38 in conjunction with Tables 34 to 37. Each of the latter gives the probability of schizophrenia and schizoidia in the children of probands of one of the four clinical subgroups of schizophrenia. In these tables columns for "definite" and "doubtful" schizophrenia appear side by side. In Table 38 all four subgroups are combined but only the figures for the definites are taken, and the total of 16.4 per cent is arrived at from them.

5. (a) Pastore has failed to grasp that the percentages for schizophrenia and schizoidia in the children of two schizophrenic parents are expectancy and not net figures. Once this is grasped, the criticism of the data yielding a figure of over 100 per cent falls away, as does the allegation of overlapping categories. (b) There is no excuse for Pastore's doubt concerning the group to which the figure of 9.1 per cent refers, as Table 60 on page 168 is clearly headed "Probability of Schizophrenia in Proband-Siblings." It can certainly not be said of this group that "errors due to partial accumulation of data are probably the largest."

6. (a) The only precise criticisms brought against Kallmann's claim of a gene-coupling between the hereditary predispositions to schizophrenia and to tuberculosis are that the schizophrenic figures included the "doubtful cases" and that the selection of the 16.4 per cent expectancy figure was made on the inadmissible grounds of choice because of the lesser complexity of the less accurate statistical method. We have already refuted both these contentions in different contexts.

(b) The allegation that Kallmann did not set forth the reasoning that led him to infer that there is a genuine gene-coupling is simply untrue. On page 246 there is a table (76) summarizing the salient numerical data, and on page 247 he sets forth his argument as follows:

1. the mortality rate from tuberculosis is highest in the descent group having the highest expectancy for schizophrenia, namely, among the proband-children;
2. the frequency rates for both disease groups in the other categories of blood-relationship decrease gradually;
3. these declining curves take an almost completely parallel course.

The correspondence between the two series of figures is so far-reaching that it entirely precludes coincidence, and even exceeds the possibility of purely numerical correlation. How extensive this agreement actually is, may be demonstrated beyond question by the calculation of the proportion between the respective frequency figures. We then see that:

4. the ratio between the figures for proband-siblings and proband-children is practically identical, both in expectancy of schizophrenia and mortality from tuberculosis, and comes to 0.7 per cent in both cases;
5. even the corresponding ratio between grandchildren and children of probands ranges only from 0.26 per cent to 0.36 per cent.

This mathematical result forges the last and most important link in the chain of our systematic taint-study in the descent of our probands. It allows hardly any other interpretation than the assumption of the *closest biological relationship between the predisposition to schizophrenia and tuberculosis*, AND above all refutes the possibility of an explanation on the grounds of similar manifestation conditions in both disease groups. At the very least, it indicates a definite gene-coupling between the tendencies to these two diseases, points to an identical pattern of heredity, and confirms the conclusion, partly established by our previous results, that both schizophrenia and tuberculosis represent *recessive* genetic factors.

(c) The calculation of the expectancy figures for schizophrenia and for mortality from tuberculosis was based on the same group of persons. It shows, moreover, a lack of understanding of scientific and statistical method to suppose that the same group of persons is necessary to demonstrate gene-coupling: witness the case of mapping the chromosomes of *Drosophila* where a number of different samples were employed.

7. (a) It is pathetic to read Pastore's statement that "a striking result in Kallmann's investigation (not brought out by Kallmann) is the finding that the offspring of probands developed only schizophrenia and no other psychosis," when a whole chapter of twenty-four closely printed pages headed "Frequency of Psychopathologic Traits Other than Schizophrenia and Schizoidia in the Descendants of the Probands" appears in Kallmann's book. Kallmann's central conclusion from the considerations adduced in this chapter is that "there are absolutely no biological or hereditary relations between the heredity circle of schizophrenia and the other abnormalities."

(b) It is my duty to draw attention to the crudity and inaccuracy of the formulation "the offspring of probands developed only schizophrenia and no other psychosis." The point at stake is whether the incidence of other mental disorders (making due allowance for selective factors) is *any higher* in the offspring of schizophrenics than in the general population.

(c) Pastore gathers that the offspring of probands develop, on the average,

the same form of schizophrenia as the parents. This statement further increases one's growing conviction that Pastore has not read *The Genetics of Schizophrenia* with understanding. Kallmann's actual conclusion is "that only about half of the schizophrenia in children and grandchildren of our probands correspond to the disease form of the related probands." The evidence upon which this is based is summarized in Table 56, and the argument against different hereditary predispositions for the various schizophrenic subgroups is found on pages 149 and 150.

(d) Pastore states that there are sex differences for such hereditary predispositions. He apparently lacks the genetic knowledge to specify whether he is postulating sex-linkage or sex-limitation. Suffice it to say that Kallmann, after adducing a vast array of numerical data, concludes:

The percentages for schizophrenia and schizoidia in all groups of our proband-children, reveal such thoroughgoing *independence of the sex* of the parental proband that it seems unnecessary to subdivide the other descent groups of our survey similarly, according to the sex of the proband (p. 109).

A more detailed statistical analysis of this point is to be found on page 124.

Summary on Pastore's Review

Even had Pastore proven all his points instead of making a series of *faux pas* through ignorance of genetic and statistical methodology, his sweeping generalization (whereby he sets at nought a thorough, systematic, genetically and psychiatrically enlightened study of such proportions that it took ten years to complete) would have been untenable: the hereditary nature of schizophrenia and even the type of genetic mechanism deduced by Kallmann would not have been upset, for Pastore emphasizes minutiae, neglecting the general effect of the vast body of Kallmann's data which he does not criticize.

I trust that this reply of mine has succeeded in showing up the fallaciousness even of Pastore's criticism of the minutiae, resulting from his lack of training in the disciplines (genetic ideology and population statistics) relevant to the type of study under consideration.

Confirmatory Evidence for Kallmann's Viewpoint

In conclusion, let me adduce confirmatory evidence for Kallmann's salient findings and viewpoint by (1) referring to his own more recent studies (twin-family) in schizophrenia and tuberculosis, and (2) making a general summary statement of conclusions from studies of other workers in this field.

1. *Kallmann's later work on schizophrenia and tuberculosis.* (a) *Schizophrenia.* In November, 1946, Kallmann published his analysis of 691 American schizophrenic twin index families (6), which confirmed

in all ways the salient findings of his German study, on which *The Genetics of Schizophrenia* was based. One of the conclusions of this paper—namely, that the genetic theory of schizophrenia “is equally compatible with the psychiatric concept that schizophrenia can be prevented as well as cured”—may appear paradoxical to those unacquainted with the concepts of modern physiological genetics. So far from equating the notions of heredity and irreversibility, this science shows how by the identification of the organic, biochemical substratum of hereditary and constitutional anomalies, a direction is given to research, which now has a tangible, more limited field upon which to concentrate. Replacing the hereditarily deficient chemical in diabetes mellitus is a case in point. Viewed in this light it will be understood how heredito-constitutional abnormalities may be regarded as holding out a better prospect of ultimate cure than psychogenic troubles, which are at the mercy of an environment which for the individual can never wholly be controlled. The heredito-constitutional view of schizophrenia has already borne fruit in the realm of practical application to therapy. Not only does it offer a rationale for the efficacy of shock treatment in terms of strengthening the functional efficiency of the reticulo-endothelial system, which is thereby better able to protect the brain against noxae (possibly endocrine in nature), but it also explains the effects of exercise in the open air and of weight regulation on the progress or arrest of schizophrenia. Kallmann has developed this theme in certain recent papers (7, 8, 9).

(b) *Tuberculosis*. Two articles (10, 11) written by Kallmann in collaboration with David Reisner, head of the Bureau of Tuberculosis, New York City Department of Health, confirm Kallmann's earlier conclusions as to a specific hereditary predisposition to tuberculosis, and elaborate the particulars of a multifactorial genetic mechanism modifying the resistance to this disease. The material investigated comprised 657 twin pairs and their families, from hospitals and clinics in the State and City of New York, reported over a period of about five years.

2. *The findings of other investigators*. The following is a summary statement of the findings of other investigators confirming the correctness of Kallmann's genetic view of schizophrenia: (a) No population survey has hitherto yielded a general expectancy rate for schizophrenia of over 1 per cent; (b) no study of a representative group of blood relations of schizophrenics has failed to yield a *significant increase* in this expectancy rate.

BIBLIOGRAPHY

1. HURST, L. A. Seventh International Congress of Genetics. *Ment. Hyg.*, N.Y., 1939, 23, 677-682.
2. HURST, L. A. Heredito-constitutional research in psychiatry. *South African med. J.*, 14, 1940, 384-400.
3. HURST, L. A. Hereditary factors in mental disorder and mental defect. *South African med. J.*, 1944, 18, 397-401.
4. KALLMANN, F. J. *The genetics of schizophrenia*. New York: Augustin, 1938.
5. KALLMANN, F. J. The scientific goal in the prevention of hereditary mental disease and race inferiority. *Proc. Seventh Int. Congr. Genet.*, Edinburgh, Aug. 23-30, 1939.
6. KALLMANN, F. J. The genetic theory of schizophrenia. An analysis of 691 schizophrenic twin index families. *Amer. J. Psychiat.*, 1946, 103, 309-322.
7. KALLMANN, F. J. Modern concepts of genetics in relation to mental health and abnormal personality development. *Psychiat. Quart.*, 1947, 21, 535-553.
8. KALLMANN, F. J. Genetics in relation to mental disorders. *J. ment. Sci.*, 1948, 94, 250-257.
9. KALLMANN, F. J. Applicability of modern genetic concepts in the management of schizophrenia. *J. Hered.*, 1948, 39, 339-344.
10. KALLMANN, F. J., & REISNER, D. Twin studies on the significance of genetic factors in tuberculosis. *Amer. Rev. Tuberculosis*, 1943, 47, 549-574.
11. KALLMANN, F. J., & REISNER, D. Twin studies on genetic variations in resistance to tuberculosis. *J. Hered.*, 1943, 34, 269-276.

Received October 12, 1950.

CORRELATION VERSUS CURVE FITTING IN RESEARCH ON ACCIDENT PRONENESS: REPLY TO MARITZ

MILTON L. BLUM AND ALEXANDER MINTZ

City College of New York

Maritz (3) states that the technique of correlating accident records in two successive periods is indispensable as evidence of accident proneness. He suggests that the fitting of theoretical Poisson and Negative Binomial distributions is not an adequate criterion of the absence or presence of differences in accident proneness in a group. There are certain weaknesses in this position, and a clarification is demanded. To reiterate the major points of our earlier paper (4):

1. Personal accident proneness, a component of accident liability, has been overemphasized.

2. This can be demonstrated by a method which reveals the extent to which accident records *could* be attributed to differences in accident liability, and this was found to be 20 per cent to 40 per cent of the total variance of accident records.

3. This method is based on the use of univariate distributions.

We agree with Maritz that a good Negative Binomial fit does not prove the existence of differences in accident proneness with mathematical certainty. We never said that it did. It is doubtful whether anything is ever proved with mathematical certainty in empirical sciences.

The issue raised by Maritz is that the correlation technique is indispensable in the establishment of accident proneness and that the evidence from a univariate distribution is invalid. We strongly disagree. We shall show that the two techniques give much the same information and, therefore, either can be offered as evidence for accident proneness. In fact, neither is wholly conclusive. Detailed histories of individuals' accident careers should be better than either.

Maritz views correlations as the "direct technique" of establishing differences in accident proneness. His claim is that "the most direct method of establishing proneness in a group of people all of whom ought to be exposed to the same environmental risk, consists of splitting a lengthy period of observation into two periods and correlating the frequency of accidents per individual for these two periods. This statistical technique is nearest to the psychological definition of accident proneness." He attempts to show that this technique may contradict conclusions based on univariate distributions. His evidence is based upon two hypothetical distributions and a previously unpublished

distribution by Adelstein. One of his hypothetical distributions resembles a chance pattern and yet yields a correlation in successive periods. The other is suggestive of differences in accident proneness but is uncorrelated in successive periods. Adelstein's data, in his opinion, illustrate two simple chance distributions which are correlated with each other.

Maritz' two hypothetical distributions illustrate the mathematical possibility for two Poisson distributions to be correlated, and for a Negative Binomial Distribution to result from summation of two uncorrelated distributions. However, the occurrence of both kinds of bivariate distributions in the case of accident records is unlikely. The mathematical derivation of the bivariate correlated Poisson distribution which Maritz applies to accident records shows only that it can be approximated by drawing colored balls from enclosed boxes. We do not believe that such a distribution should be expected in the case of accident records. The only obvious meaning of a Poisson distribution of accident records is that it is satisfactorily explained by the assumption of equal and constant accident liability. Without such an assumption it looks like a result of an odd coincidence. If Poisson distributions of accidents are the results of equal liability, they should be uncorrelated, and it is not clear what kind of combination of circumstances should lead to the expectation of Poisson distributions correlated in successive periods.

In his other hypothetical distribution this lack of plausibility is obvious. For example, all six hypothetical individuals who had more than eighteen accidents each in one period had zero accidents in the other period. This could occur, but is hardly to be expected. In other words, in this example, Maritz proves that it is possible to obtain a correlation of $-.11$ by arranging numbers in a manner designed to obtain it. This is granted, but what does it prove about accidents?

The only empirical material Maritz presents is the unpublished Adelstein data, and it must be regarded with more seriousness than his two hypothetical distributions, which are mathematically possible but highly improbable. Maritz claims that the examination of the univariate distributions suggests a "pure chance pattern," but that the correlation between the accident records in the two periods reveals the existence of differences in accident proneness to the extent of a correlation of $.29$. The existence of the correlation is treated as something that could not have been foreseen in terms of the accident distributions in the two observation periods. Thus, the impression is created that the combined period had properties which were essentially different from

those of the shorter periods when considered alone. He bases his interpretation on the fact that three χ^2 tests fail to reveal significant differences either between two Poisson distributions and Adelstein's five- and six-year distributions, or between a bivariate correlated Poisson distribution and Adelstein's scatter diagram for the two periods.

In dealing with these data Maritz makes the common error of confusing the failure to disprove a hypothesis with its proof. He states: "Equation [1] was fitted to the observed data of Table III and the resulting test of goodness of fit gave for 7 *df*, $P = .49$. Hence it follows that the above data follow a *correlated* bivariate Poisson distribution" (p. 438). This second sentence does not follow from the first. The failure to disprove a hypothesis according to which the data for a combined period have properties (the .29 correlation) different from those of the constituent periods (viewed as exhibiting a simple chance pattern) is not the same as proof of it. A closer examination of the data for direct evidence of this type of heterogeneity fails to reveal anything convincing, and the opposite and theoretically more plausible hypothesis of essential homogeneity of the eleven-year observation period fits the data more closely than Maritz' heterogeneity (i.e., bivariate correlated Poisson) hypothesis.

In terms of the properties of the accident distributions in the two consecutive periods, the most probable correlation between the accident records in these periods is not zero as Maritz implies, but .21. This is quite close to the observed .29. The estimate of a correlation of .21 was arrived at by determining the estimated percentages of the accident records attributable to factors other than chance (18 per cent and 24.4 per cent) and then computing their geometric mean. In determining these percentages the formula $(v-m)/v$ was used; the two variances were 1.485 and 1.370, respectively, and the two means were both 1.123.

If the eleven-year period had no properties essentially different from those of the shorter period, it should be possible to construct theoretical scatter diagrams approximating the empirical one by utilizing statistics derived either from the two periods considered separately (without considering their correlation), or from either one of them taken alone, or from the total period taken as a whole. The form of the bivariate distribution chosen was the bivariate Negative Binomial. It is based on much the same assumptions as those made by Greenwood and Yule (1) in their derivation of a univariate unequal liability distribution, namely:

Accident liability is distributed in people in accordance with a Pearson III curve.

Accident liability of a person remains constant per unit of time throughout the two observation periods.

Each particular degree of accident liability gives a simple chance (Poisson) distribution of accident records in each observation period (p. 279).

Table I presents the Adelstein data together with a theoretical distribution constructed in accordance with these assumptions; the computation utilizes only the mean and the variance of the first period, and the fact that the second period lasted six years while the first one lasted five years. The formula for the theoretical frequency for the cell representing j accidents in the first, k accidents in the second is

$$y_{jk} = Na^b \left(\frac{c}{c+a+1} \right)^p \frac{\Gamma(p+j+k)}{\Gamma(p)j!k!(c+a+1)^{j+k}}$$

in which a is the ratio of the two durations, and c and p are two constants derived from the mean and variance of the first period, as follows: $c = m/(v-m)$, $p = m^2/(v-m)$. (The derivation of the formula, which closely resembles that of Greenwood and Yule, will be published elsewhere.)

TABLE I

A COMPARISON OF ADELSTEIN'S ACTUAL ACCIDENT DATA WITH THE THEORETICAL BIVARIATE DISTRIBUTION COMPUTED FROM HIS FIRST-PERIOD DATA
($m=1.123$, $v=1.370$)

First Period	Second Period							
	0	1	2	3	4	5	6	7
0	21(16.3)	14(14.8)	8(8.0)	1(3.3)	—(1.2)	—(0.4)	—(0.1)	
1	17(12.3)	12(13.4)	8(8.4)	3(4.1)	1(1.6)	—(0.6)	—(0.2)	1(0.1)
2	6(5.6)	9(7.0)	2(5.1)	2(2.7)	2(1.2)	—(0.5)	—(0.2)	—(0.1)
3	1(1.9)	1(2.8)	3(2.3)	3(1.4)	1(0.7)	—(0.3)	—(0.1)	
4	1(0.6)	3(1.0)	—(0.9)	—(0.6)	—(0.3)	—(0.1)		
5	—(0.2)	—(0.3)	—(0.3)	2(0.2)	—(0.1)			
6	—(0.1)	—(0.1)	—(0.1)	—(0.1)				
Totals	46(37.0)	39(39.4)	21(25.1)	11(12.4)	4(5.1)	0(1.9)	0(0.6)	1(0.2)

The theoretical distribution appears to fit the data quite well. The χ^2 computed was 8.743, $df=10$, $P=.58$. It follows that the correlation technique recommended by Maritz did not add anything significantly new to the information one could gather by examining one of Adelstein's univariate distributions.

In arguing for correlations and against properties of univariate distributions as grounds for assuming differences in accident proneness,

Maritz overlooks the fact that these two kinds of statistical measures are closely interrelated. The correlation between accident records in two periods can always be computed from the variances in these two periods and in the total period, according to the formula

$$r = \frac{V_{1+2} - V_1 - V_2}{2\sqrt{V_1 V_2}}.$$

Similarly, the increase in accident variance when one combines two observation periods is an increasing function of the correlation between these periods, in accordance with the elementary formula $V_{1+2} = V_1 + V_2 + 2r\sqrt{V_1 V_2}$. Inasmuch as this formula is exact, and inasmuch as every observation period has an early stage when the variance is smaller than the mean, a distribution cannot have a variance greater than the mean value unless there were positive correlations between successive periods somewhere in the past. Similarly, a Poisson distribution cannot result unless the accident records in the subdivisions of the observation period were uncorrelated or unless the effects of positive and negative correlations cancelled each other. If an examination of two univariate distributions in two successive periods suggests something markedly different about the existence of accident proneness, compared to the correlation between these periods, the interperiod correlations must have similarly changed in the past. Maritz' hypothetical distributions could be used just as readily in arguing against the use of correlations in accident research as against the use of variances. The only advantage of a correlation lies in the fact that it enables one to tell a particular time when the variance of a set of accident records has risen at a rate beyond chance expectation.

A further consideration with reference to the correlational technique is that it presents certain practical difficulties. Accident proneness is a problem for industry as well as for the theoretical statistician. From the point of view of industry it is often difficult to acquire data on the same individual for two successive periods. Those with high accident rates in the first period are likely not to be found in the second period. They may be dismissed or resign from their jobs, not to mention being hospitalized or dead. The study reported by Kerr (2) is typical of practical problems confronting industry. The major effort is to reduce accidents, not to wait for successive periods.

SUMMARY

1. The hypothetical distributions presented by Maritz are mathematically possible and demonstrate the lack of mathematical certainty

of inference from empirical data, but such distributions are not likely to be encountered in practice.

2. Correlational research on accident proneness is legitimate, but inferences about accident proneness drawn from correlations are not more certain than inferences drawn from the fitting of distributions.

3. Correlational research is not always feasible for practical reasons. In any event, it is not indispensable with reference to establishing accident proneness.

BIBLIOGRAPHY

1. GREENWOOD, M., & YULE, G. U. An enquiry into the nature of frequency distributions representative of multiple happenings, with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *J. roy. statist. Soc.*, 1920, 83, 255-279.
2. KERR, W. A. Accident proneness of factory departments. *J. appl. Psychol.*, 1950, 34, 167-170.
3. MARITZ, J. S. On the validity of inferences drawn from the fitting of Poisson and Negative Binomial distributions to observed accident data. *Psychol. Bull.*, 1950, 47, 434-443.
4. MINTZ, A., & BLUM, M. L. An examination of the accident proneness concept. *J. appl. Psychol.*, 1949, 33, 195-211.

Received November 18, 1950.

CONCERNING TASTE-BLINDNESS TO PTC

JOZEF COHEN

University of Illinois

AND

DONALD P. OGDON

University of Texas

A recent paper by Boyd (1) has offered several criticisms of a review paper on taste-blindness to PTC published by Cohen and Ogdon (2). This note will reply to those criticisms, using the same paragraph numbering as Boyd's paper. References will not be repeated since they may be found in the bibliographies of the above communications.

1. In the first paragraph of his paper, Boyd has said that our report that Lasselle and Williams were the first discoverers of taste-blindness (to creatine) is false. Boyd indicates that he investigated this substance shortly after publication by Lasselle and Williams and found that no taste-blindness existed. The trouble with this sort of affair is, that all things being equal, we believe what is reported in the literature. We read what Lasselle and Williams had found, and we included it in our review. We did not read what Boyd had found for the reason *that he did not put it into the literature*. Since we didn't read it, we didn't mention it.

2. In this section Boyd refers to a "careful study" by Hartmann as finding that non-tasters could not taste PTC even when it was dissolved in saliva from strong tasters. This study, says Boyd, should have been included in our review.

Miss Hartmann did not come to the above conclusion. She found, in fact, that one non-taster (herself) could not taste PTC when it was dissolved in the saliva of one taster. The *entire* study is reported in a single sentence; there is not the slightest mention of concentration, technique, nor is there the slightest mention of a control. The total *N* was equal to 1. We consider the description of the experiment as a "careful study" as being ludicrous, and reject the suggestion that it should have been included in our section concerning the role of saliva in the tasting of PTC.

3. Here Boyd discusses sex differences in taste-blindness to PTC. Boyd says that we should have included three studies in support of his claim that sex differences are significant. These studies are Boyd and Boyd, Falconer, and Riddell and Wybar. We specifically cited the results of Boyd and Boyd in our paper. We had not seen the Falconer

paper at the time we wrote the review. The reason is that the review was written in 1948, and the Falconer paper was published in 1947 in England. It takes about a year for material printed in England to be brought to America and abstracted by the appropriate agencies. It had not come to our attention.

The suggestion that we should have included the Riddell and Wybar study in support of sex differences in taste blindness is something we call to the attention of the careful scientist. Riddell and Wybar state clearly and precisely that their data *show no significant differences*.

4. Boyd discusses here the racial distribution of tasters with reference to our table containing, by far, the majority of studies completed in this area. We maintain two things: (a) the numbers do not form racial clusters, (b) the table is inconsistent. Racial clusters are not obvious, and we have yet to see a map drawn with accurate contour lines. The table is inconsistent because American Caucasians have values all the way from 60 per cent to 82 per cent. We did not say that there are not significant differences between groups indicated in the table. Significant differences do occur, probably even between different experiments on what are supposed to be the same group—American Caucasians.

We are glad to learn that Boyd has developed a standard technique for testing for taste to PTC. His paper describing the more important aspects of it was published *after* our review appeared; we were unable to include the technique because we were writing about PTC and not clairvoyance. In any event, other investigators did not use Boyd's technique, and their results, as we have said, are not directly comparable.

5. This paragraph, recapitulating our reports of taste-blindness to di-phenyl-guanidine and thiouracil, is based on a 1950 paper by Boyd and an unpublished manuscript by Boyd and Hoffman, and we will not comment on it. A review paper, or a criticism of a review paper, should not contain any new material, and we object to criticisms involving experiments which were unpublished at the time of publication of the review paper.

BIBLIOGRAPHY

1. BOYD, W. C. "Taste blindness" to phenylthiocarbamide and related compounds. *Psychol. Bull.*, 1951, 48, 71-74. blindness to phenyl-thio-carbamide and related compounds. *Psychol. Bull.*, 1949, 46, 490-498.
2. COHEN, J., & OGDON, D. P. Taste- Received March 27, 1951.

REPLY TO TRAVERS' "A CRITICAL REVIEW OF THE VALIDITY AND RATIONALE OF THE FORCED-CHOICE TECHNIQUE"¹

DONALD E. BAIER²

Department of the Army

The Personnel Research Section of The Adjutant General's Office is an operating research agency. It conducts both basic and applied research for the purpose of providing the Army with the best possible personnel tools. As an operating research agency, it sometimes, because of the pressure of events, delays publication of results of interest to the psychological profession. By the time research is published, all too often progress has been made toward further refinement or, sometimes, even toward a different viewpoint on a given problem. The recent review by R. M. W. Travers of the forced-choice technique (23) is a case in point.

Travers is in the unfortunate position of attempting a review of a problem which originated in the Personnel Research Section and on which but a small fraction of the Section's research has been published in the psychological journals. However, reports of additional research are available as Personnel Research Section Reports³ and as papers read at meetings of the American Psychological Association. This additional information would have kept Travers from making such statements as the following (23):

1. The following quotation summarizes all the statistical evidence that the present writer has been able to obtain concerning the relative validity of the forced-choice (FCL) and the more traditional type of rating device (RCL) as it is used in the Army Officer Efficiency Report (p. 67).
2. The claims for the validity of the technique seem to bear little relationship to the actual evidence (p. 66).
3. Proper studies need to be made to determine the validity of scales in this area constructed on the basis of an adequate rationale (p. 70).

Travers could have obtained the additional information had he but indicated his intent to write a review. The difficulties of generalizing in the field of personnel research, resulting from the operation of so

¹ TRAVERS, R. M. W. A critical review of the validity and rationale of the forced-choice technique. *Psychol. Bull.*, 1951, 48, 62-70.

² The opinions expressed herein are those of the author and do not necessarily reflect the views of the Department of the Army.

³ Personnel Research Section Reports are not available for general distribution. Arrangements have been made to furnish the American Documentation Institute with copies of unclassified reports for distribution.

many factors, are well known. A critical review is scarcely worth the name if it leans heavily on a single study and does not cover other research information.

There is one highly important omission from Travers' article. His title and his remarks are directed at the forced-choice technique in general, although the data he discusses are confined to rating procedures. It should not be overlooked that the forced-choice technique has been used in other types of instruments, for example, personality inventories and self-description forms. The forced-choice technique has demonstrated its great usefulness in the construction of such instruments (12, 13, 16, 19). Even higher validities have been obtained with modifications of conventional forced-choice technique based on suppressor theory (1, 20). The lack of explicitness in Travers' review on this point is misleading. Compared to the traditional "yes-no" type of questionnaire, the technique has produced personality measures of useful predictive value for the Army situation. Working with the traditional type of personality items has consistently failed to yield a useful product.

The comments in this reply refer only to rating scales used for efficiency-reporting or merit-rating purposes as involved in the work of the Personnel Research Section.⁴ Further, the paper is not intended as a complete review of the research concerning this application of forced-choice. It is an effort to discuss the problems raised by Travers, and to make available some of the more general findings concerning the problem of efficiency reporting. This effort will be made in terms of: (1) areas in which there is agreement with Travers, (2) areas in which there is disagreement with him, and (3) areas or problems concerning which he makes no comment.

AREAS OF AGREEMENT

Ratings should not be validated against other ratings (23, p. 69). From one point of view there is no doubt of the desirability of "criteria other than judgments." The Personnel Research Section and other investigators have searched and still are searching for more objective and appropriate criteria. Practicable suggestions for the development of such criteria would be eagerly welcomed. Until such a development occurs, investigators will no doubt continue to use ratings as criteria, and considerable effort will be expended toward improving such ratings.

The problem of criteria for efficiency reports deserves more than the

⁴ The problems involved in securing ratings for criterion purposes are not necessarily the same.

brief comment Travers gives it. In considering the use of ratings versus objective criteria in any instance, the nature of "success" being predicted must be carefully considered. In some instances, ratings may be the best criteria because value judgments are the essential elements. What must be clearly recognized also are the problems of interpretation that are involved when ratings are used to "validate" ratings. More specifically, three points should be noted concerning the use of a composite of ratings as a criterion.

1. It is considerably better than no criterion at all. It is well known that averaging a series of ratings will tend to reduce bias. At the very least, therefore, use of multiple ratings as criteria in evaluating rating scales will improve rating procedures by identifying those scales which contain the least amount of bias.

2. Use of composite ratings as a criterion would seem to have its maximum justification in studies of performance as an officer. This is the case, since performance as an officer involves, as a large component, the ability to work with and through other people. Furthermore, an officer's career involves a large variety of duty assignments; the expression of his value must be in generalized terms. Judgments of superiors, subordinates, and immediate associates are especially pertinent.

3. Use of multiple ratings as criteria creates problems of interpretation of the findings involving comparison of specific rating techniques. Up to the present, the rating composite has essentially been an averaging of ratings obtained by a single technique—the traditional type of rating scale. When a rating scale is involved as a predictor, one never knows the extent to which it is favored because of its similarity to the criterion. Indirect evidence suggests that the amount of such "technique contamination" is appreciable (17). The solution to the problem of comparing rating techniques, when criteria differing entirely in character are unavailable, may be the inclusion of all types of rating techniques in the criteria. This procedure will give each rating technique an equal opportunity of showing "validity." Such a procedure reduces the problem of "validity" in rating studies to one of rater agreement, i.e., reliability, if this concept is considered to cover a relationship of one rater using a given technique to several raters using the same technique.

In the sense of choosing between two members of a pair, forcing a choice is not an essential part of the technique. In discussing the rationale of the technique, Travers makes a great deal of the point that all items of a pair or a tetrad could either be listed in rank order or the rating could be given in terms of a traditional rating scale with the restriction that no two traits could be rated at the same point (23, pp. 64-65). This may well be true. The possibility has already been indicated in conjunction with self-rating items (7, p. 186). Inasmuch as it is obviously difficult, if not impossible, to extend indefinitely the number of items which can be considered together, an element of forcing is bound to be present. This is nothing new. Choices must be made among the terms

that are grouped together, much in the same sense that a choice must be made among the alternatives of any multiple-choice item. This point is relatively unimportant except for its relationship to the next.

Forced-choice pairs work because the nonscored alternative serves as a suppressor. This is an important point because of its theoretical significance. We agree that suppressor theory may provide the rationale for the success of forced-choice items; in fact, we have exploited it heavily in connection with self-description inventories, as mentioned above (20). To avoid any misunderstanding, however, certain points should be made explicit:

1. Travers states that forced-choice procedure assumes "for any given individual, the true rating on the irrelevant [i.e., unscored] items is average and that any deviation in the average ratings on these irrelevant characteristics represents a tendency to over-rate or under-rate the particular individual who is being rated The procedure is simply that of using ratings on certain characteristics as a suppressor variable to correct for errors in the rating on certain other characteristics" (23, p. 65).⁵

2. The above assumption is not necessary, nor was it made in the development of the forced-choice procedure. To quote from one of the Section's early papers (11) on this technique, "The essence of the forced-choice technique, as we use the term, however, is the grouping of the alternatives to make them appear of equal value, and yet have unequal significance." In other words, items are paired so as to give each alternative equal face validity and differing true validity. Whether or not individuals have an average rating on "irrelevant" items is not essential for either of these conditions to obtain, nor has it any bearing on suppressor theory in forced-choice items.

3. The suppressor theory requires only that (a) the scored alternative of a pair have as high a validity as possible; and (b) the nonscored alternatives have as low (even negative) a validity as is consistent with a high relationship with the scored item.

4. A casual reader of Travers' article might conclude that a separate suppressor key could be developed for the traditional rating scale items. Travers did not suggest this possibility, but it is desirable to make explicit the point that this application of suppressor theory may not work. Use of traditional ratings as suppressors for other traditional-type ratings has been tried in obtaining rating criteria. The suppressor theory has been confirmed in the sense that negative Beta weights were obtained for the intended suppressor ratings, but the effect was so slight that validity of the combined ratings was not improved (18). It is not intended to assert that a suppressor key might not be developed on the basis of traditional-type items, but only that available evidence does not encourage optimism in this belief.

Forced-choice items can be improved in their content (23, pp. 63-64).

⁵ "Irrelevant" and "relevant" are perhaps not sufficiently meaningful in this context. In one sense, all items are "relevant" to officer performance. "Discriminative" or "nondiscriminative," "differentiative" or "nondifferentiative," "critical" or "non-critical" are suggested substitutes.

Pairing items on a statistical basis only will frequently bring together alternatives which normally would not be associated. Indeed, this may be one of the advantages of the forced-choice procedure. Travers' point that the content of the pairs should not confuse the rater is well taken. However, we believe he has exaggerated the problem. Directions to the rater have always stressed that he is to indicate which alternative *most nearly* applies to the person he is rating. Items such as Travers cites do meet the crucial test of having and *maintaining validity* over a period of time (8).

Forced-choice items do not prevent the rater from manipulating his rating if he so desires (23, pp. 69-70). In publications of the Personnel Research Section, claims have been much more modest than Travers implies. To cite one instance, "... it reduces the rater's ability to produce any desired outcome of obviously good or obviously bad traits. It, thus, diminishes the effect of favoritism and personal bias" (10). The emphasis is on the words "reduces" and "diminishes."

Personal bias is a general term indicating departure from the true value for any reason. Bias may result from insufficient information on which to base a rating, from the unconscious operation of friendship, from differences in leniency on the part of raters, etc. It is in the reduction of these types of bias that the forced-choice technique may be particularly helpful. The rater who deliberately desires to manipulate his rating can undoubtedly do so. However, the forced-choice technique makes it a somewhat more difficult task for him.

In passing, it might be pointed out that an efficiency report is primarily a means of recording the rater's estimates. By itself, regardless of technique used, it does not guarantee that the rater will be honest, comprehensive, careful, and objective. To achieve this purpose, supplementary aids must be used, and even these may not be effective. In the Army, this aid is in the form of an Army Regulation which contains not only the necessary administrative procedures but also a discussion of the purpose and use of the efficiency report and of the psychological principles involved in rating. This psychological information would not have been included in the Regulation if it were believed that the forced-choice technique were an automatic and complete control of rater bias.

In relation to this question, it should be pointed out that while a rater can move his rating up or down the traditional-type rating scale at will, and can influence the score he is giving on forced-choice items, on neither type of rating scale can he determine with much precision the relative standing of the person he is rating. This point is most clearly seen when scores on rating scales are translated into some

standard scale. It is not uncommon on a seven- or eight-point rating scale for 30 per cent of the responses to be concentrated at a single point. The amount of change a swing of one point on a scale will produce on a standard score is evident. Unless, therefore, a rater knows precisely the distribution of ratings, he can never know where he is placing a person on a relative population scale, the kind of rating used by the Army. This point is mentioned because it is believed that a good deal of the objection encountered by the forced-choice technique has been misdirected, and the point to which objection is taken is basically the difficulty of reconciling relative and absolute standards.

AREAS OF DISAGREEMENT

Some of the areas of disagreement are quite minor. These will be disposed of first.

1. *Travers is incorrect in his statement, "Each one of these elements describes, in essence, a rather specific item of behavior" (23, p. 62).* A glance at the alternatives shows very many general terms, i.e., *modest, no one ever doubts his ability, low efficiency, businesslike*. One of the unsolved problems is the degree of specificity which alternatives in forced-choice groupings should possess.

2. *Travers is incorrect in his interpretation of preference index.* He states, "The other value [preference index] indicates the extent to which individuals tend to rate others too high or too low on a particular characteristic" (23, p. 62). The preference index is, to quote from an early publication on this technique (11), an index of the "value to the rater" of the alternative under consideration; more recently, the preference index has been considered as a measure of the face validity of the item. It is hoped that use of the forced-choice technique will tend to correct for raters' tendency to rate too high or too low, but this is not involved in the computation of the preference index.

We disagree with Travers' statement, "Claims for the validity of the technique seem to bear little relationship to the actual evidence" (23, p. 66). In support of this statement, Travers relies heavily on a minor study of the Personnel Research Section (14) and one by Richardson (6). As mentioned earlier, in making this statement he has ignored the vast body of research data available. Some of the data have been presented at recent meetings of the American Psychological Association (2, 4, 8, 9, 21).

It would take us too far afield to review this work here. Perhaps the information in Table I (from 15), based on two samples totaling 7,771 cases, will suffice to indicate the type of evidence available to support the statement that "... it [a combination of forced-choice and graphic rating scales embodied in an official efficiency report] produces ratings which are more valid indices of real worth" (10). This table reports

TABLE I

COMPARATIVE VALIDITY OF FORM 67 AND FORM 67-1, APRIL, 1946, EDITION (FROM 15)

Rank	Sample 1 (N=4,208)		Sample 2 (N=3,563)	
	Form 67	Form 67-1	Form 67	Form 67-1
Col.	.24	.35	.30	.30
Lt. Col.	.13	.23	.48	.50
Maj.	.32	.42	.32	.34
Capt.	.21	.31	.34	.35
1st Lt.	.34	.46	.45	.51
2nd Lt.	.30	.45	.46	.57

some of the results of a study conducted in connection with the regular reporting period of 30 June 1946. Both WD AGO Form 67 and WD AGO Form 67-1 were completed for the same officers. The score on Form 67 was an average of ten eight-point graphic rating scales; the score on Form 67-1 was a combination of forced-choice and rating scales. The criterion used was an average of ratings by superiors, subordinates, and associates obtained by a nominating technique. The consistently greater validity of Form 67-1 is evident.

We disagree with Travers' interpretation of the quotation "[the forced-choice rating technique is] relatively free from the usual pile-up at the top of the scale" (23, p. 66). In the first place, Travers has confused the forced-choice technique per se with Form 67-1. This form contains both forced-choice and traditional-type rating scales. The distributions he reproduces (from 10) are for the total score on Form 67-1. Travers does not observe this distinction; hence, his remarks are misdirected.

In the second place, Travers does not comment on the difficulties in comparing a distribution based on a scale of 220 used points (Form 67-1) with a distribution based on a scale of 43 used points (Form 67).⁶ The attempt to equate the range for the purpose of comparing distributions on the two forms gives the traditional rating scale (Form 67) every advantage.

In the third place, Travers has missed the point. A Personnel Research Section Report, dated 17 January 1947 (15) contains the information on which were based the illustrative Distributions reproduced as his Figure 1. The computation of the third and fourth

⁶ The score on Form 67 has a possible range of -4 to +7. Scores below 2.7 are rare. Considering the score in tenths of a point gives 43 points in the actual range.

moments contained in that report shows that Form 67 has greater leptokurtosis and that Form 67-1 has greater negative skew. "This will mean that Form 67-1 will be more discriminative of extreme cases than will 67, particularly at the low end of the distribution" (15, p. 7).

In the same Personnel Research Section Report, there appear data (15, p. 16) which show the percentage of officers at two cut-points on the distributions. These data are reproduced as Table II. If the equated data are taken at face value, this table clearly brings out the better discrimination by Form 67-1 at both ends of the distribution. Below the point where the curves (23, Fig. 1) cross at the low end of the distribution, there are 18.2 per cent for Form 67-1 and 9.5 per cent for

TABLE II
PERCENTAGE OF OFFICERS BEYOND LOW AND HIGH CUT-
POINTS ON FORM 67 AND FORM 67-1
(From 15, Table V)

Grade	% to Lower Cut from Bottom			% to Upper Cut from Top		
	Form 67	Form 67-1	% Excess	Form 67	Form 67-1	% Excess
Colonels	31.46	60.28	28.82	No cross	No cross	0.00
Lt. Colonels	24.69	48.15	23.46	No cross	No cross	0.00
Majors	13.28	33.65	20.37	3.32	4.53	1.21
Captains	12.06	20.24	8.18	5.07	10.10	5.03
1st Lts.	3.56	12.93	9.37	10.32	17.22	6.90
2nd Lts.	5.51	13.53	8.02	9.72	17.89	8.17
Combined	9.48	18.16	8.68	8.55	10.67	2.12

Form 67. Above the point where the curves cross at the high end of the distribution, there are 10.7 per cent for Form 67-1 and 8.6 per cent for Form 67. Table II shows the same type of information by grade. The point of particular interest is that for the lower grades, the difference in effectiveness of the two forms is most pronounced at the high end of the scale; for the upper grades, the difference is most marked at the lower end of the scale. In the light of this kind of data, there is no question as to which form is the more useful.

We disagree with Travers' conclusion: "The data [from 14] suggest that experimentation with different types of directions may yield much more important results than experimentation with forced-choice scales" (23, p. 68). Experience of the Personnel Research Section indicates that this statement represents an extreme oversimplification of the problem involved. In our experience, the basic attitudes of the raters, determined in large part by knowledge of the uses to which a rating scale

is to be put, are little influenced by specific directions.⁷ If psychological effects of such a kind influence ratings, it would appear more likely that they will be brought about by the rater's actually doing something (for example, some form of preliminary ratings) than by a certain type of instruction. Furthermore, raters do not necessarily follow the instructions. Attention is called to the fact that despite what must be an extraordinary variety of directions for merit rating forms, high negative skew and leptokurtosis are almost invariably characteristic of the ultimate distributions. These stubborn characteristics, in fact, have served as motivation for the search for other than the traditional rating techniques.

We disagree with Travers' interpretation of a rater agreement represented by a correlation of 0.69 (23, p. 69). While it is not the intention to discuss Travers' comments on Richardson's study, from which this "reliability" coefficient is cited, it should be pointed out that the "only 0.69" leads to the wrong evaluation of a coefficient of this magnitude. This may be illustrated from a follow-up study of the validity of Form 67-1 (17). In this study, rater agreement on the criterion ratings obtained at a single sitting was represented by $r = .24$. Agreement between the official Form 67-1 raters for 914 cases was represented by $r = .56$. In Army experience at least, rater agreement as represented by coefficients as high as .70 is a rare finding and not to be considered unusually low.⁸

EFFICIENCY-REPORTING PROBLEMS OMITTED FROM TRAVERS' ARTICLE

In attempting a critical review of a technique, it is customary to discuss problems which it was hoped this technique might solve, and to cite the complete evidence. These comments would seem to be particularly pertinent to articles appearing in the *Psychological Bulletin*. We have already indicated deficiencies in the citation of evidence. It will, perhaps, clarify the problem if a short history is given here.

It is a truism in rating-literature that halo, leniency, and rater differences in standards are basic problems. The Army Officer Efficiency reporting system, as exemplified in WD AGO Form 67, had become increasingly subject to these influences.⁹ Figure 1 illustrates the

⁷ The evidence for grade bias, i.e., higher ranking officers being rated higher, in Table II, despite careful instructions to disregard grade, is a case in point.

⁸ Before leaving this section, a slight error should be corrected. Travers attributes the quotation beginning "a single over-all rating on a 20-point scale . . ." on page 67, to his reference No. 4. The quotation is actually contained in his reference No. 5.

⁹ It is probably more accurate to say that Form 67 had become increasingly subject to halo and leniency. No direct evidence is available concerning variations in rater agreement.

increasing tendency for officers to be rated higher with the passage of time. Form 67 was not believed by the Army to be serving its purpose, largely because it had lost its discriminating value at the high end of the scale. The Army directed the Personnel Research Section to develop a rating system which would meet its needs to a greater degree. The basis for the research on the problem of efficiency reporting was the study concerned with the development of procedures for the integration of officers into the Regular Army following World War II. This research

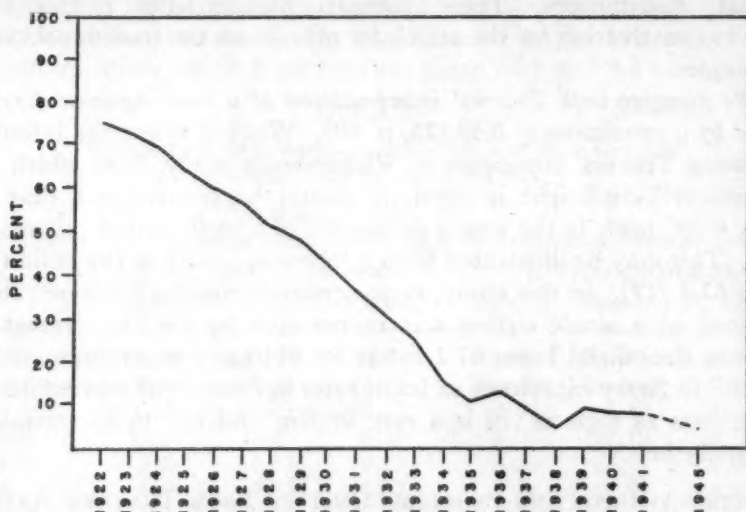


FIG. 1. PERCENTAGE OF ALL REGULAR ARMY CAPTAINS RECEIVING LESS THAN EXCELLENT RATINGS ON WD AGO FORM 67, 1922-1945
No Ratings for War Years 1942-1944.

has been outlined by Richardson (5). The first Personnel Research Section studies of the forced-choice technique were undertaken under this program.¹⁰

¹⁰ In the interest of historical accuracy, some elaboration of Travers' statement concerning the origin of the forced-choice technique should be made. The idea of forced-choice was suggested by Dr. Paul Horst in a discussion at an APA meeting. He himself, does not remember the incident. Dr. Wherry was sufficiently interested to develop the scaling methods to achieve the process as Dr. Horst had discussed it; namely, to present, simultaneously, items which looked alike to the individual completing a personality scale and yet had differing significances. Dr. Wherry developed the scaling procedures while working for the Civil Aeronautics Authority and brought them with him when he came to the Personnel Research Section. Jurgensen (3) during this period had been working on a somewhat similar idea, a fact which was not known to the Personnel Research Section until after World War II. It was Wherry's basic scaling technique

Following this integration program, studies were initiated which compared five different rating forms, including rankings, various kinds of traditional-type ratings, and forced-choice. The first series of studies involving some ten thousand officers in the United States and Europe showed a slight superiority for the forced-choice rating form. This superiority, coupled with the hypothesis that there would be less change in the distribution for the form involving the forced-choice items when the form was used on an official basis, led to the decision to use it along with the old Form 67 at the June, 1946, rating period. Details with respect to skew and means are presented elsewhere (22). Results bearing on the validity have already been presented in Table I.

This is not the place to review in detail the further findings of these and subsequent studies. It seems more helpful at this point to summarize the advantages and disadvantages of the forced-choice procedure as applied in the Army efficiency reports.

The principal disadvantage is that the use of the technique had tended to be unacceptable to Army officers (although, apparently, more acceptable in industry). Acceptability is an especially important problem in rating procedures because of the effect on the rater's attitude. Two comments may be made about the unacceptability:

1. The name, "forced-choice," is an unfortunate one. Reasons for its origin are readily understood in the light of the fact that the original presentation asked the individual to pick one of the two items as most descriptive of himself. Actually, as previously suggested, forced-choice might better be considered a scaled multiple-choice item.

2. The second point that should be made concerning the acceptability of the items arises out of the conversion of the raw scores on efficiency reports to standard scores. Converting to a relative scale caused raters to feel that their ratings were not properly represented by a particular standard score, especially those below average. In objecting to Form 67-1, there was much confusion between the effects of the forced-choice technique and the effects of the use of a relative standard score scale.

The advantages of the technique may be summarized as follows:

1. It reduces halo. Raters completing graphic ratings within the same form tend to mark them all pretty much the same way; i.e., the correlation between graphic rating scales is high. In completing two sections of forced-choice items, raters likewise tend to mark them pretty much the same; i.e., forced-choice sections also correlate high, but not as high as rating scales. Forced-choice ratings and rating scales correlate less than do rating scales with rating

which served as a point of departure for the work of the Personnel Research Section, first in the application to the development of personality inventories and later in the application to efficiency reporting. As noted in this article, the technique has been most successful in application to self-rating, i.e., personality inventories.

scales, or forced-choice with forced-choice ratings (17). On the simple basis of lower intercorrelation, a combination of the two techniques would have greater possibility of increasing validity. In comparison with Form 67, Form 67-1 has persistently yielded slightly greater validity, perhaps for this reason.

2. It reduces bias; for example, it is less influenced by rank of the rater-officer than was Form 67 (10). On the whole, the total score on Form 67-1 agrees better with an average criterion rating than do any of its sections (9, 17). The use of the average in itself is a conventional means of reducing bias.

3. Forced-choice item validities tend to be stable over a period of time. From December, 1946, to January, 1949, item validities correlated from .50 to .60 (8). This stability is especially noteworthy in view of the narrow range of these item validity coefficients. Although this evidence needs confirmation, it is the sort of evidence which encourages experimentation with the technique.

4. Raters agree better on a report composed of both types of technique than they do on either type alone (17). This finding is of first importance. Ratings on Form 67-1 for two successive rating periods for a group of 914 raters showed the best agreement for a combination of both techniques (i.e., total score) than for either technique (Table III).

TABLE III
AGREEMENT OF RATERS ON SUCCESSIVE REPORTS, FORM 67-1

<i>Traditional rating sections:</i>		<i>Forced-choice sections:</i>			
	<i>r</i>		<i>r</i>		
Section V.....	.47	Section IV.....	.42	<i>Total Score</i>56
Section VII.....	.39	Section VI.....	.45		

In the absence of other than a rating criterion, the problem of validity of efficiency reports may reduce to one of this kind of reliability. Thus, the greater rater agreement on a combination of the two techniques is of special significance.

It should be noted that the criterion used for the validity coefficients presented above was the average of a series of rating scales. Rating scales would, therefore, be favored, a point which Travers neglects to mention in his discussion. A further point which is illustrated in the above tabulation is that rating scales and forced-choice ratings may differ among themselves. It is, therefore, difficult to make any hard and fast generalizations concerning either type of rating.

The research of the Personnel Research Section, plus certain theoretical considerations, has persistently affirmed the value of the forced-choice technique. Since in the Army, at least, efficiency reports are usually considered for the entire career and since no technique or combination of techniques has brought rater agreement up to a satisfactory value, our attention has been directed to developing a system whereby fluctuations in rating owing to leniency or other purely biasing

factors might be reduced. To put it another way, so much more is gained by combining ratings made by different raters than by improving the rating of a single rater through the use of a special technique that our emphasis is on averaging the reports prior to making use of them. Obviously, for such a system to work, an adequate distribution of single ratings must be maintained. From Figure 1 it is clear that over a period of time, ratings on the traditional type of rating scale in the Army tend to become restricted to the upper portion of the scale. If this kind of trend can be established as characteristic, it would appear to be necessary to develop techniques such as forced-choice which have the promise of maintaining a spread in the ratings.

It should be further noted that in the industrial situation, where people may be rated consistently by the same rater, this averaging system probably will not work. In such situations, therefore, an effort to develop techniques such as forced-choice is particularly needed.

In conclusion, three points should be emphasized: Travers discusses forced-choice technique as applied by itself. It should be observed that both in this reply and in the research reports of the Section, the value of forced-choice in combination with the traditional type of rating scales has been stressed. Until an experiment is set up which gives each technique an equal chance to prove its worth—that is, the correlation not being subject to technique contamination—conclusions as to the value of techniques used will not be definitive.

Secondly, the forced-choice technique has been discussed in terms of the way it has been applied. There are many ways in which it might be improved, e.g., in method of calculating preference and discriminative values, in method of pairing alternatives, perhaps by further application of suppressor theory or better grouping in terms of item content. And, finally, the well-demonstrated value of the technique in the construction of self-rating scales of the personality inventory or self-descriptive type is again stressed.

BIBLIOGRAPHY

1. BROGDEN, H. E., NEWKIRK, G. F., & LOEFFLER, J. C. The prediction of officer potential of ROTC cadets. *Amer. Psychologist*, 1950, 5, 360. (Abstract.)
2. CAMPBELL, J. T., & RUNDQUIST, E. A. Scale items for inclusion in forced-choice rating form. *Amer. Psychologist*, 1950, 5, 280. (Abstract.)
3. JURGENSEN, C. E. Report on the "Classification Inventory," a personality test for industrial use. *J. appl. Psychol.*, 1944, 28, 445-460.
4. KARCHER, E. K., & KING, S. H. Effect of number and order of ratings on reliability and validity. *Amer. Psychologist*, 1950, 5, 333. (Abstract.)
5. RICHARDSON, M. W. Selection of Army officers. In *New methods in applied psychology*, Proceedings, Maryland Conference on Military

- Psychology. Univ. of Maryland, 1947, 79-85.
6. RICHARDSON, M. W. An empirical study of the forced-choice performance report. *Amer. Psychologist*, 1949, 4, 278-279. (Abstract.)
 7. RUNDQUIST, E. A. Personality tests and prediction. In D. H. Fryer and E. R. Henry, *Handbook of applied psychology*. New York: Rinehart, 1950. Vol. 1, 182-191.
 8. RUNDQUIST, E. A., WINER, B. J., & FALK, G. H. Follow-up validation of forced-choice items of the Army Officer Efficiency Report. *Amer. Psychologist*, 1950, 5, 359. (Abstract.)
 9. SCHNEIDER, DOROTHY E., & BLACKBURN, J. R. Validity of a graphic scale of officer efficiency. *Amer. Psychologist*, 1950, 5, 359. (Abstract.)
 10. SISSON, E. D. Forced-choice—the new Army rating. *Personnel Psychol.*, 1948, 1, 365-381.
 11. STAFF, PERSONNEL RESEARCH SECTION, AGO. The forced-choice technique and rating scales. *Amer. Psychologist*, 1946, 1, 267. (Abstract.)
 12. STAFF, PERSONNEL RESEARCH SECTION, AGO. Construction and selection of items for the Biographical Information Blank (BIB). PRS Report No. 703, 7 July 1945.
 13. STAFF, PERSONNEL RESEARCH SECTION, AGO. Validation of Form E of the Biographical Information Blank. PRS Report No. 716, 8 July 1946.
 14. STAFF, PERSONNEL RESEARCH SECTION, AGO. Comparison of the rating check list (RCL) and forced-choice list (FCL) methods of obtaining ratings. PRS Report No. 717, 9 July 1946.
 15. STAFF, PERSONNEL RESEARCH SECTION, AGO. A comparison of officer efficiency ratings obtained with the WD AGO Form 67 and the WD AGO Form 67-1 at the Regular Reporting Period of 30 June 1946. PRS Report No. 725, 17 Jan. 1947.
 16. STAFF, PERSONNEL RESEARCH SECTION, AGO. Item analysis and development of scoring keys for the Leaders' Course BIB. PRS Report No. 764, 1948.
 17. STAFF, PERSONNEL RESEARCH SECTION, AGO. 1949 Follow-up validation of WD AGO Form 67-1. PRS Report No. 791. (In preparation.)
 18. STAFF, PERSONNEL RESEARCH SECTION, AGO. Determination of preference and discrimination values of phrases for preference check list sections of alternate forms of Officer Efficiency Report, WD AGO Form 67-1. PRS Report No. 846. (In preparation.)
 19. STAFF, PERSONNEL RESEARCH SECTION, AGO. Development and validation of the Self-Description Blank for predicting leadership qualities of ROTC cadets. PRS Report No. 859. (In preparation.)
 20. STAFF, PERSONNEL RESEARCH SECTION, AGO. A rationale for minimizing distortion in personality questionnaire keys. PRS Report No. 868. (In preparation.)
 21. TAYLOR, E. K., CARROLL, J. B., & WINER, B. J. Validity of the Army's officer efficiency report. *Amer. Psychologist*, 1949, 4, 284. (Abstract.)
 22. TAYLOR, E. K., & WHERRY, R. W. A study of leniency in two rating systems. *Personnel Psychol.*, 1951, 4, 39-47.
 23. TRAVERS, R. M. W. Validity and rationale of forced-choice technique. *Psychol. Bull.*, 1951, 48, 62-70.

Received April 19, 1951.

NOTE ON TRAVERS' CRITICAL REVIEW OF THE FORCED-CHOICE TECHNIQUE¹

MARION W. RICHARDSON

Richardson, Bellows, Henry and Co., Inc., New York

With reference to the hesitation of Travers "to accept the evidence provided by that author [Richardson] and his associates since the procedure involved seemed to raise spuriously correlations between assessments of job performance based on a forced-choice scale and an independent criterion of job proficiency," Richardson pleads *mea culpa* in only one respect. The reviewer did not have access to enough data, because of the slowness of publication. That is our fault.

Once the question came up as to the evaluation of psychological research performed for industrial clients by agencies such as Richardson, Bellows, Henry and Co., and my suggestion was that any research results reported by us would be referred willingly by us to a competent competitor such as the Psychological Corporation. I am confident that the Psychological Corporation would have similar reactions. In this situation, we should give Dr. Travers the opportunity to examine the basic evidence both in the raw and in the analyzed form. It does not necessarily follow that because new techniques are being tried out in a nonacademic environment, the procedures are unsound.

I will try to suggest some of the things the reviewer will discover if he accepts the invitation or when RBH properly gets around to publication.

First, let us dispose of the purified criterion issue by some generalizations no more broad than the reviewer's.

1. *The purified criterion technique is not a necessary feature of the forced-choice technique.* It has not always been used in the thirty odd performance reports developed by our firm.

2. *The technique probably has the effect of reducing the dimensionality of the behavioral complex we call "effectiveness on the job."* If true, the effect is to simplify the measurement problem by eliminating some of the variance due to the more elusive factors. Surely it is legitimate to decide whether one does the simpler job reliably or the more complicated one unreliably.

3. *If one applies the scale to the entire sample population (thus using the unpurified criterion), the shrinkage in the correlation coefficient is not so large as Travers imagines.*

In some of our studies the purified criterion is used to select the

¹ TRAVERS, R. M. W. A critical review of the validity and rationale of the forced-choice technique. *Psychol. Bull.*, 1951, 48, 62-70.

original items for inclusion in the groups (blocks of three, four, or five), and, again, to select blocks from the total number tried out. (The latter procedure is not mentioned by the reviewer, since Sisson did not list it.) Although the validity coefficient may be computed on the purified criterion, it has always, or nearly always, been recomputed on the entire sample in a manner that ought to satisfy the statistical purist's soul. In our studies the shrinkage has not been alarming, but Travers has a right to see the evidence in forthcoming publications, or otherwise.

The real statistical issue is the shrinkage of validity coefficients on successive samples drawn from the same population. This is the *bête noir* of most validity studies. The one scale criticized by the reviewer was applied by a competent independent investigator to a new group not used for standardization of the scale.

M. Trawick of the Esso Standard Oil Company reports on the same scale, a validity coefficient of $r=0.76$. A key official of the Standard Oil (N.J.) Company has (as yet unpublished) *résumés* of several repeat studies on the scale criticized by Travers. With the purified criterion issue completely absent, and on a background of two years' experience with the scales under rigorous tryouts in new situations, the validity coefficients ranged from $r=0.60$ upward, and several exceeded (within limits of standard error) the original coefficient reported. Under the conditions set, there was no possibility of criterion contamination.

We are much more interested in such tests, some of them on new samples with mixed cultural and linguistic groups, than we are on a minor statistical point such as the reviewer brought up. We can at least retort "guilty in part, but what of it?"

This amount of correlation shrinkage has not, to date, worried our investigators. We do not know whether the use of the purified criterion technique increases, leaves unchanged, or decreases the all-important shrinkage on new samples. It may be argued on an arm-chair, but inconclusive, basis that the *important* shrinkage is actually decreased, but more evidence is needed.

The present writer is inclined to agree that further evidence on negative skew and elimination of control of final rating by the rater is needed.

If the criticized may turn on the critic, I would express the opinion that the "rationale" portions of the article are too fragmentary and incomplete. A proper rationale would be mathematical, involving a "difference" function the present writer has not yet been able to handle.

The reviewer's one constructive suggestion seems to involve directions too complicated to be carried out on ordinary people, and in its

feasible special case, to lead us right back to the technique he doesn't like. For example, how much difference is there between *forcing* raters to give a different rating to every trait, and *forcing* a choice of one of two or more traits as most characteristic?

Again, the writer pleads *mea culpa* only with respect to his failure and that of his associates to furnish data on the now-numerous applications of the so-called forced-choice techniques. (There is no "stock-in-trade forced-choice instrument," as averred by the reviewer.)

Received April 19, 1951.

BOOK REVIEWS

MOWRER, O. HOBART. *Learning theory and personality dynamics*. New York: Ronald, 1950. Pp. xviii+776. \$7.50.

Mowrer has taken a leading role over the last few years in providing integration at the level of theory between the two fields suggested by the book's title: the laboratory studies of learning, and the clinical study of the person. The struggle upon which he has been engaged calls for a reinterpretation and reconciliation of the viewpoints stemming from Freud and Pavlov.

The book represents selections from the author's earlier papers, supplemented by a good deal of previously unpublished material. Following an introductory chapter, the book is divided into two parts. The first part, entitled "Learning Theory," consists of twelve chapters, of which three contain previously unpublished material. The second part, "Personality Dynamics," also twelve chapters, includes four new ones. The previously published chapters are supplemented by footnotes to indicate the author's changes in point of view since the papers first appeared.

It will be convenient to permit remarks about the book to correspond to the two parts. First we shall consider the contributions to learning theory, followed by attention to what the author has to say about personality.

Learning theory. The two major contributions to learning theory are, first, the conception of acquired fear as a secondary drive, and second, a dual theory of learning, distinguishing between conditioning (simple associative learning) and problem-solving (learning with reinforcement, under the law of effect).

That fear-reduction may be reinforcing is now quite widely accepted by learning theorists whose points of view are related to those of Hull. Mowrer deserves credit for originally proposing this conception, and for building the apparatus (with Miller) that has served to provide convenient laboratory demonstrations of the possibility of reinforcement through escape from noxious stimuli, and, in later modifications, through escape from fear.

Accepting fear as a learned drive is not the same as accepting fear as the *only* learned drive, although Mowrer at points comes near to doing this. Impressed by the role of fear in learning through punishment, he finds that some experiments on food or water deprivation may also be interpreted through the mediation of fear. Such interpretations are offered of an experiment by O'Kelly and Heyer (p. 347) and one of Estes (p. 349).

This is not the place to review these interpretations in detail, but the reader has to be warned that the treatments are not entirely self-

consistent, and such inconsistency as remains is not solely a matter of the range of dates covered by the original papers. Mowrer's greatest difficulty comes in trying to explain secondary reinforcement by a stimulus associated with reward. A familiar example is the reward-value of the click associated with the food-delivering mechanism in a Skinner Box. In a passage inconsistent with his fear-reduction theory, Mowrer has this to say of the secondary reinforcing effect of a signal produced by the organism's response: "However, what is important here is merely that the signal take on a *pleasant connotation*, so that when the subject happens to produce the signal, the response which was effective in producing it will be reinforced" (p. 323).

Psychoanalytic theory gives a central place to anxiety-reducing mechanisms. The familiar defense mechanisms are said to serve this purpose. To show the operation of related principles at the rat-learning level has provided a useful bridge between animal learning and psychoanalytic theory. This much we may accept, without accepting fear as the only learned drive, and recognizing with Mowrer that the rat's fear is not the same as the human being's anxiety. We owe Mowrer a debt for initiating and advancing this line of thinking.

The second feature of learning theory, stressed over and over again in the book, is the need for a dual theory. Mowrer has rather reluctantly joined the ranks of the many earlier writers who made a distinction in kind between simple associative learning and rewarded learning. He attempted earlier, with Hull, to reduce all learning to the paradigm of reinforcement. But now he is convinced that the original effort was faulty. His footnotes, reinterpreting earlier results, and many supplementary discussions, seek to make a case for the modified position.

Conditioning (the first of the two kinds of learning) is "the process whereby emotional learning . . . takes place" (p. 236). "It now seems preferable to apply the term 'conditioning' to that and only that type of learning whereby *emotional* (visceral and vascular) responses are acquired . . . Many responses which have previously been termed 'conditioned responses,' are in the present conceptual scheme, not conditioned responses at all. Only those responses which involve visceral and vascular tissue and are experienced subjectively as emotion are assumed to be conditioned responses" (p. 244).

Problem-solving (the second of the two kinds of learning), is equated with effect learning. "Effect learning has been previously conceived as applying mainly to those situations in which the motive, or 'problem,' is an unlearned biological drive, such as hunger, thirst, pain, etc. It is now clear that effect learning may be expanded to include those situations in which the motive, or 'problem,' is a *learned* drive, i.e., an emotion such as fear or an appetite. . . . If an emotion, or secondary drive causes the skeletal musculature to be activated, and if such activity results in secondary drive reduction, then the overt response

thus acquired is here conceived as an instance of effect learning, not conditioning" (p. 244).

The problem that gave rise to this reinterpretation was that of how the acquired fear came about that was to serve as the secondary drive, and hence provide reinforcement when reduced. Mowrer has decided that the solution is to accept a simple association between the fear responses and the incidental stimuli present at the time that fear responses occur. Hence a box in which an animal is shocked (and frightened) will later produce fear even in the absence of shock. This fear serves as a drive to motivate learning that will reduce it.

The choice that Mowrer felt called upon to make (and that he invites his reader to make) is between a monistic reinforcement theory and his dual theory. There are many more alternatives. Mowrer makes a pretty strong case against a monistic reinforcement theory, but this by no means leaves his dual theory as the only (or even the preferred) alternative.

Others who accept simple contiguity theories (e.g., Thorndike's associative shifting, and Guthrie's conditioning) do not find it necessary to limit such associative learning to emotional responses. The evidence for so limiting associative learning is no better than for extending it, say, to word-association learning.

The criticisms of the monistic reinforcement theories are to be taken seriously. The dual theory that Mowrer proposes has to be examined as one of the alternatives, but it is not the only one, and needs much tightening to become a strong candidate for wide acceptance.

Personality dynamics. One is led to expect a much clearer integration between learning theory and personality dynamics than he actually finds.

The duality theory of learning is said to have its counterpart in Freud's distinction between the reality principle and the pleasure principle. To be sure, the pleasure-principle, concerned as it is with immediate gratification and tension-reduction, sounds like the law of effect. But does this, then, make the reality principle correspond to primitive emotional conditioning? In Freudian theory, the reality principle comes in later on, as the organism learns to postpone gratification in order to deny the painful consequences of seeking immediate pleasure. This is problem-solving; it is not simple conditioning. Mowrer might have toyed with Freud's repetition principle as coming nearer to simple associative habit. But this possible analogy he did not follow.

The treatment of personality dynamics arises much more out of reflection upon Freud's theory as it applies to clinical data than out of inferences based on learning theory. The main emphasis is upon disagreements with Freud. This is interesting because Mowrer was one of those who, in earlier writings, helped make Freudian theory plausible

to experimentally oriented psychologists. The somewhat reluctant disagreement with Freud parallels in a way the disagreement with reinforcement theory.

As specimens of the disagreement we may consider the theory of identification, the theory of neurosis, and the theory of anxiety.

Mowrer believes that both boy and girl achieve an identification with the mother, before sex-roles are differentiated. The sex-role identification is then with the like-sexed parent, coming later. This is contrasted with the Freudian theory of a first choice on a sexual basis ("object-choice") of the opposite-sexed parent, followed later by identification.

Mowrer accepts an immaturity theory of neurosis, as contrasted with an overlearning theory that he attributes to Freud. That is, neurosis for Mowrer signifies an underlearning of what is needed to get along, while neurosis to Freud often means the need to unlearn the prohibitions that were too well learned in childhood. It seems somewhat doubtful that this is fair to Freud, and the strong possibility exists that both kinds of neuroses may be found. In Freudian terminology, some cases have an underdeveloped superego ("holes in the superego"), while others have an overwhelming superego.

The third difference with Freud lies in Mowrer's "guilt" theory of anxiety as contrasted with Freud's "impulse" theory. In Mowrer's words:

"In essence, Freud's theory holds that anxiety comes from evil wishes, from acts which the individual would commit if he dared. *The alternative view here proposed is that anxiety comes, not from acts which the individual would commit but dares not, but from acts which he has committed but wishes that he had not. It is, in other words, a 'guilt theory' of anxiety rather than an 'impulse theory' "* (p. 537).

There are consequences for therapy arising out of these differences with Freud. In general, the relief of the pressure of guilt through confession leads to a kind of conversion experience. Neurotic anxiety becomes normal anxiety, and new learning can occur. Therapy does not consist in watering down the superego, but in growing up to the assumption of responsibility. "Valid treatment should . . . lie in the direction of helping the individual to grow up, emotionally and socially, to the point where the demands of conscience and community are understandable and acceptable" (p. 572).

We have here a very provocative book, certainly alive to the issues that are being debated in contemporary psychology. The book is difficult, but it is thoughtful, and the careful reader can learn a great deal from it.

It is to be regretted that the author did not make the effort to rewrite his papers instead of merely annotating them. The volume as it is offered the reader is more difficult than it needed to be, for it carries

along all the asides and preoccupations of the author as he was feeling his way toward his final theory. A much smaller book might have carried his message more clearly to a wider audience.

Stanford University.
University of Michigan.

ERNEST R. HILGARD.
EDWARD L. WALKER.

VÖLGYESI, FRANZ ANDREAS. *Hypnosetherapie und psychosomatische Probleme*. (Hypnotherapy and psychosomatic problems.) Stuttgart: Hippokrates-Verlag Marquardt & Cie., 1950. Pp. 203. DM 8.25.

There would be little purpose in reviewing a book that is very unlikely to come to the attention of most American psychologists and psychiatrists, were it not for the fact that, at least according to the publishers (Germany, U.S. Zone), the author's pronouncements aroused great interest among the readers of a German medical journal (*Hippokrates*), and, far more important, that the book gives a very significant glimpse of the "progress" of psychological science behind the Iron Curtain. The ominous suffusion of scientific work with political ideology of which this book—probably against the author's intentions—gives eloquent evidence, is sufficient reason for Western readers to take a good look at the product and shudder at the perversions of science under a dictatorship: from there it is but a short step to the abominations perpetrated by the Nazis, when calculated sadism and utter disregard for human dignity paraded under the guise of "research."

Völgyesi is a Hungarian medical researcher who has written extensively on the subject of hypnotherapy and reflexology, deriving his theoretical inspiration from Pavlov, Bekterev, and their students, whose work is cited throughout the book. Reflexology, then, is the conceptual framework for the author's hypnotherapy, the aim of which is "a general, central reorientation of the patient by means of psychocortical, cortico-visceral, psychosomatic, suggestive effects, which are determined by conditioned reflexes" (Foreword, p. 7). This becomes clearer in chapter II, which is entitled "So-Called 'Psychosomatic Medicine.'" It is admitted there that "Anglo-Saxon" psychosomatic medicine has had a fruitful influence, but certain "essential corrections" have to be made. These encompass (1) the historical side of the psychosomatic viewpoint, (2) the psychoanalytic overemphasis, and (3) the "so-called 'independent psychosomatic diseases.'" The main line of attack is directed against the psychoanalytic orientation of psychosomatic medicine in this country, and the "favorite attempt" to postulate "independent psychosomatic syndromes." We learn that the latter do not exist, just as there is no "purely somatic" or "purely mental" illness. In every disease, both aspects are said to be inextricably inter-

woven. Quite apart from the fact that the foregoing is a gross misrepresentation of present-day thinking in psychosomatic medicine, it is more instructive to be told that since the time of Aristotle "*body and soul have formed a social-dialectic, that is, antinomical entity.*" (All italics are the author's.)

A great many examples are cited to demonstrate that "*active hypno-suggestive psychotherapy achieves its results despite somatic, toxic, etc. disturbances*" (p. 147). And, quoting from Grastchenkov, "*the relationships between etiology and pathogenesis cannot possibly be treated scientifically without the principles of reflexology*" (p. 102). Asserting that all neuroses and all types of neurotic manifestations always have organic bases, Völgyesi attempts to establish "dialectically interpreted" hypnotherapy as a technique which supposedly produces astounding cures for a variety of medical and psychiatric problems. But this is not all; not only is hypnotherapy *the* psychotherapy par excellence (psychoanalysis is alternately termed "reactionary" and an "aberration"), but it also produces structural changes in the organism's repertoire of unconditioned reflexes. Thus, it somehow affects the hereditary characteristics of the organism. On page 91 we read: "Suggestion and conditioned reflexes are capable of altering, with experimental exactness, the hereditary unconditioned reflexes."

If these were one man's views, they could easily be disregarded. But the multitude of experimental investigations cited show with incontrovertible clarity that this represents "official" Soviet thinking:

The final summary of the joint serial sessions of the Scientific Academy and the Medical Scientific Academy of the Soviet Union, in June 1950, made serious reproaches (*erhob schwere Vorwürfe*) against Orbeli, Beritasvili, Anochin, Kupalov, even against Speransky, and their schools: despite all of their valuable achievements relative to conditioned reflexology, they did not follow sufficiently *the original concepts and the methodology of Pavlov's doctrines* regarding nervism and materialistic neuro-psychiatry, and the concluding note of the final summary "*calls upon every worker in physiology and medical science, to promote in constructive fashion, with free criticism and self-criticism, the great doctrine of Pavlov, for the benefit of the people*" (footnote 5, p. 104).

Implicit in these formulations are (1) an assumption concerning an almost limitless modifiability of the human organism, which makes it capable of being directly influenced by suggestion, (2) a paradoxical rigidity once modifications have been made through suggestion, such that the hereditary structure is permanently altered, and (3) a fanatical belief in the all-inclusiveness of conditioned reflex principles as an explanatory construct. Irrespective of the scientific truth of the theory, which is at best partial and certainly contravenes much evidence available today, this book is a glaring example of what happens when research is built around political ideology and when experimental findings are interpreted after the scientist has made certain that they

fit in with "official" thinking. May this never come to pass in American psychology!

HANS H. STRUPP.

*Human Resources Research Laboratories,
Bolling Air Force Base.*

ROSENBLUETH, ARTURO. *The transmission of nerve impulses at neuro-effector junctions and peripheral synapses.* New York: Wiley, 1950. Pp. xiv+325. \$6.00.

Part I of this monograph stems from the 1937 work by Cannon and Rosenblueth, *Autonomic Neuro-Effector Systems*. It is concerned with the liberation of chemical mediators by nerve endings at their junctions with smooth muscle, cardiac muscle, and glands. It reviews the evidence that acetylcholine is the substance liberated by the cholinergic axons located chiefly in the parasympathetic system. The controversial problem of the adrenergic axons is dealt with by supposing that either adrenaline or an "adrenaline-like substance" is released at their endings, which occur chiefly in the sympathetic system. The action of the adrenaline-like substance is further supposed to be conditional upon the presence of a receptive substance at the junction with the effector; it is the excitatory or inhibitory nature of this substance which results in the formation by adrenaline of one of two hypothetical substances, sympathin E and sympathin I. Rosenblueth thus maintains in this work the principal contention of Cannon and Rosenblueth in their 1937 monograph: that the adrenergic effects are those of a mediator combined with a local agent whose function is to regulate the excitatory or inhibitory response of any particular effector at any given time. While there is some discussion of recent experiments in that area, the present treatment contains no change in basic point of view.

Part II has to do with peripheral synapses, broadly conceived to include not only the neuro-neuronal junctions in autonomic ganglia but also the junctions of somatic motor nerve fibers with striated muscle. The author states, "The argument for chemical transmission at peripheral synapses appears stronger than that which can be made for chemical transmission at autonomic neuroeffector junctions, yet the latter transmission is generally accepted as chemical whereas the former is still considered electrical by many of the experts in the field." Rosenblueth cites as evidence in support of chemical transmission experiments on Wedensky inhibition, Wallerian degeneration, and the Philipeaux-Vulpian phenomenon. He considers the effects of curare, veratrine, and potassium ions, and assigns to the latter the function of adjuvant to acetylcholine, the principal mediator. The spike potential is regarded as a mere manifestation of a depolarization of nerve membrane which is assumed to be the condition essential for the release of acetylcholine.

Incidentally, there is no mention in the book of the ideas of Dr. Norbert Wiener, whose book, *Cybernetics*, contains a tribute to Rosenblueth for his understanding of scientific methodology.

However well presented, the thesis of this book will no doubt be widely challenged, and Rosenblueth will be criticized for clinging too tenaciously to theories advanced more than a dozen years ago by Cannon and himself. This he does in spite of recent significant work of Bacq, Bozler, Gaddum, Feldberg and others. Rosenblueth does take cognizance of this newer work, notably in chapter IV, but it is apparent that it has had little effect in modifying his interpretation of the role of electrical events, to cite a specific example, in smooth muscle contraction. J. H. Gaddum some time ago remarked drily that there could be no doubt of the importance of the experiments by Cannon, Rosenblueth, and others which led to the theory of sympathins E and I, but that the theory itself should be forgotten. It has been Gaddum's contention that no known substance has the properties ascribed to sympathins E and I, whereas all the experimental observations can be accounted for on the basis of known substances, notably adrenaline and noradrenaline.

Another authority in this field, Z. M. Bacq, is of the opinion that acetylcholine, adrenaline, and noradrenaline are the "local hormones" or chemical mediators which are clearly established at the present time. He states that there is universal agreement among English, Swedish, and Belgian pharmacologists on the point that sympathin is a mixture, in variable proportions, of adrenaline and noradrenaline. The confident enthusiasm of Rosenblueth may be contrasted with some recent observations of G. L. Brown to the effect that we are perhaps within sight of an explanation of the mode of action of acetylcholine, and we are in a similar state of obscurity with respect to the action of adrenaline as a chemical mediator.

The reviewer, no expert in these matters, has talked with Dr. F. G. Sherman of Brown University and other colleagues among the physiologists. He has received the impression that Rosenblueth's monograph will be received by serious students as an excellent compilation and introduction to the difficult literature on chemical mediators. Yet there is a distinct tendency to go along with the authorities enumerated above in rejecting hypothetical mediators in favor of those whose properties and chemical structure, at least *in vitro*, are well defined.

LORRIN A. RIGGS.

Brown University.

GOULDNER, ALVIN W. (Ed.) *Studies in leadership*. New York: Harper, 1950. Pp. xvi+736. \$5.00.

Studies in Leadership is a collection of thirty-two papers representing the work of thirty-five different authors. Twelve of the selections have

been published previously, while twenty are new contributions. After an introduction by Gouldner the papers are organized into five groups. Part one, "Types of Leaders," deals with informal leaders, bureaucrats, and agitators. "Leadership and Its Group Setting" is the title of the second part, which is concerned with leadership in various social classes, in several minority groups, and in diversified political atmospheres. Part three, bearing the provocative title "Authoritarian and Democratic Leaders," deals with problems of democratic, manipulatory, and authoritarian leadership. Part four is called "The Ethics and Techniques of Leadership," and the descriptive title of the final group is "Affirmations and Resolutions."

Something of the tone of the volume can be gathered from an examination of the headings given above and even more from a consideration of the method and scientific orientation of the articles. According to the reviewer's classification, four of the papers are largely historical. Typical of such articles is Cox's "Leadership among Negroes in the United States" or Green and Melnick's article on the feminist movement. Seven contributions are oriented around material gathered through the field observation of some leader or social group. Some of Leighton's important conclusions, based on material gathered in the relocation camp for Japanese at Poston, are presented. There is an interesting case study of a local union leader by Alexander and Berger. Also included are three short selections from Whyte's *Street Corner Society*. Feuer presents some stimulating observations on the leadership and organization of collective settlements in Israel. Another seven studies are more "data oriented" in the sense that considerable statistical material is given as a basis for the presentation. Typical here is material from "The People's Choice." Particularly interesting is a study by Lipset on the leadership of the Cooperative Commonwealth Federation (CCF) of Saskatchewan. Finally, there is a group of fourteen contributions, or almost half the articles, which are predominantly speculative, philosophical, or theoretical. While a number of the essays in this group are stimulating, provocative, and insightful, many of them are not scientific in the sense that the authors do not make use of any reasonably explicit body of factual material to serve as a basis for their theorizing.

If Gouldner's value judgments of the good in leadership work are typical of sociologists, and if these papers fairly represent such values, then a considerable gap in opinion regarding desirable methodology exists between those working in sociology and those in psychology. The reviewer feels that most psychologists are wedded to an approach to leadership problems (and social psychological problems generally) which is "fact oriented." While theory is needed, it is not allowed to run into the broadest generalizations unchecked by data. One of the authors in this collection states, "Almost the entire literature on leadership stems

in large measure from the writings of Aristotle and Machiavelli" (p. 396). Indeed, here would seem to be the major criticism of many of the selections: while interesting, they are essentially social philosophy, which, if this book is evidence, is not based on any comprehensive set of data.

Such considerations raise the question as to the selection of material. In such a fast-growing field as leadership it is difficult to keep abreast of all the newest developments, but it is disappointing not to find more contributions from recent psychologically-oriented leadership research. There is nothing from the Ohio State group (Shartle, Stogdill, Hemphill). Cattell's work on syntality is not mentioned. The new developments in leadership and group behavior associated with the Research Center for Group Dynamics is inadequately portrayed by one of Lewin's early (1939) papers. (Incidentally, two of Lewin's papers are reprinted only in part and are given new titles, but neither change is indicated by the editor.) The wartime developments in leadership and group behavior of the Office of Strategic Services and of the British are represented in an article by Eaton, which has been outdated by *The Assessment of Men* and by Harris' *The Group Approach to Leadership-Testing*. Because of such omissions or inadequate representation, many psychologists will feel that *Studies in Leadership* does not give an adequate picture of current work in this field.

LAUNOR F. CARTER.

University of Rochester.

CANTRIL, HADLEY. (Ed.) *Public opinion 1935-1946*. (Prepared by Mildred Strunk.) Princeton, N. J.: Princeton Univ. Press, 1951. Pp. lix +1191. \$25.00.

Not only social scientists, but also government officials, business executives, and labor leaders have many occasions for inquiring as to the state of public opinion on a given topic. While this need is likely to be most acute as regards contemporary issues, the significance of today's poll can best be evaluated if it is placed in the context of past observations. The practical man is likely to be particularly concerned with trends: Are Americans becoming more isolationist? Is the public more hostile to labor unions?

All of us who are working with such questions will owe a real debt to Mrs. Mildred Strunk, who did most of the work in compiling this impressive reference volume. She has gathered results from 23 polling organizations in 16 countries for the 11 years from 1935 to 1946, and organized them in an easily consulted format. Evaluation is facilitated by indicating type of sample, percentage of "don't know" responses, and other significant information.

It is to be hoped that Dr. Cantril's organization will carry out its

plan for issuing such volumes at five-year intervals. Such a series would have great value for social psychology all over the world.

ROSS STAGNER.

University of Illinois.

BRITT, STEUART HENDERSON. *Selected readings in social psychology*. New York: Rinehart, 1950. Pp. xvi+507. \$2.00.

The maturity of social psychology—like that of any science—will be marked by more systematic treatment of the kind which provides empirically testable hypotheses. While the beginnings of such maturity are dimly seen in some recent books, we are far from any really satisfactory theory-building and probably will be for some time to come. In the meantime social psychology seems to be in for another spate of symposia and books of readings. In fact, there appear to be certain recurrent fashions in these matters. Perhaps in a rapidly growing and still vaguely bounded field we may expect this cyclic phenomenon.

The vague and still unstructured nature of systematic social psychology is evident in Britt's book—as it is in contemporary textbooks generally—in the variation in major topics presented as well as in the content of the individual papers. Britt has organized his selections into five major divisions: (I) "Social Psychology and Its Methods"; (II) "Biological and Social Foundations of Behavior"; (III) "Some Individual Factors of Social Adjustment"; (IV) "Behavior in the Presence of Others"; (V) "The Social Psychology of Institutions"; and (VI) "Social Conflicts."

As is bound to be true of all such books of readings, the decisions of a particular compiler as to what to include or exclude seldom completely agree with the judgments of others. On the whole this reviewer feels that Britt has done a reasonably good job, though his materials are chiefly of contemporary character. Certainly some of the papers are of such ephemeral value that they will quickly go out of date. Personally I should like to have seen some of the classical, though admittedly theoretical, papers included. For example, pertinent selections from James, Cooley, Dewey, George H. Mead, and Lippmann would provide the student-reader with an historical and somewhat systematic background.

In a review of this character one cannot summarize or criticize the whole range of the fifty selections which make up this book. In keeping with present trends Britt recognizes the importance of social-cultural materials for social psychology. Papers by Linton, Warner, La Barre, and Withers testify to this fact. The attention given to the topics of social class and social conflict, as is true in the widely used *Readings* by Newcomb and Hartley, is evidence of a growing awareness of the importance of these matters in our everyday life. So far, however, no one has come up with any very testable theory about the function of these factors in relation either to personality or group structure. Since the

editor of this volume is professionally much concerned with problems of mass communication, I should have expected more extensive and somewhat different selections on this topic. In this connection a most serious and obvious gap is the failure to cover the field of language adequately. The short selection from Albig is good as far as it goes, but the inclusion of papers on the symbolic process as related to communication, cross-cultural materials (such as from Lee and Whorf), and some of the recent publications of Merton, Lazarsfeld, and their co-workers would have strengthened the volume.

But to this reviewer there is a more serious omission. The usefulness of books of this kind, usually designed for orientation and information to the beginning student, is greatly enhanced when the editor uses interpretative summaries (or introductions) as a method of tying the materials together. Moreover, when such a job is well done, it may also serve to contribute to the growth of systematic theory and thus advance our field at least a little bit. Aside from Britt's own opening paper on methods, originally published in 1937, no effort is made to give any systematic interpretation to the field.

KIMBALL YOUNG.

Northwestern University.

HOMANS, GEORGE C. *The human group*. New York: Harcourt, Brace, 1950. Pp. xxviii+484. \$6.00.

Students of social behavior, whatever the label on their academic credentials, should be interested in this provocative book on the face-to-face or primary group. Vigorously positivistic and empirical though neither experimental nor quantitative in his approach, Homans develops in connection with a series of case studies a minimum array of concepts and hypotheses for the analysis of groups as emergent and self-regulatory systems. Indeed, his major contribution seems to me to lie in spelling out what is implicit in this conception of the group as a system, in terms of explicit relationships among analytically distinguished variables. His approach is rooted particularly in the work of L. J. Henderson, Elton Mayo, Chester Barnard, and Conrad Arensberg.

Activity, interaction, and sentiment are the three primary elements of social behavior with which Homans starts his analysis. While he adds a fourth basic element, *norms*, the mutually interdependent relations among the first three provide the organizing framework for his analysis of group phenomena. In what appears to me as a useful convention, he distinguishes between the *external system* of the group—"the state of these elements and of their interrelations, so far as it constitutes a solution of the problem: How shall the group survive in its environment?"—and the *internal system*—"the elaboration of group behavior that simultaneously arises out of the external system and reacts upon it." In the external system, the problem of motivation involves the rela-

tions between sentiment and activity, while functional specialization and the division of labor involve the mutual dependence of activity and interaction. Each of the possible pairings of variables is considered in the analysis of the internal system. Here are examples of the kinds of hypotheses he proposes under each heading:

Other things equal, contact breeds liking and vice versa. (Interaction and sentiment.)

Persons who feel sentiments of liking for one another will express those sentiments in activities over and above the activities of the external system, and these activities may further strengthen the sentiments of liking. (Activity and sentiment.)

Activities in addition to those of the external system tend to arise among persons who interact regularly, and in turn tend to strengthen their tendency to interact with one another. (Interaction and activity.)

With this conceptual equipment Homans treats in turn clique differentiation, social ranking, leadership, interpersonal relations, social control, and social change.

Five major case studies, each first presented with a minimum of conceptualization but selected to bring out some major aspect of his system, are a notable feature of the book. Faithful to the primary sources, the presentation is a far cry from the perfunctory illustration more frequent in textbooks. The Bank Wiring Room (from Roethlisberger and Dickson's *Management and the Worker*) is used to develop the broad outlines of the internal and external systems. Leadership is considered in relation to the Norton Street Gang (from Whyte's *Street Corner Society*), while the network of interpersonal relations in a family and kinship system is studied with material from Firth's *We the Tikopia*. "Hilltown," a New England community, provides a case of social disintegration (drawn from an unpublished thesis by D. L. Hatch), while a study of industrial morale in the "Electrical Equipment Company" (by C. M. Arensberg and D. MacGregor) serves as a focus for the consideration of social conflict.

I was much struck by the agility with which Homans manages to analyze the relations between particular aspects of behavior in the group without falsifying their interdependence with the entire context of the group as a system. I think social psychologists can study with profit his approach to system, "function," and equilibrium in groups. But his analytical variables seem to me too crude to comprise many of the sensible things he has to say. For example, in discussing leadership he hypothesizes that "*when two persons interact with one another, the more frequently one of the two originates interaction for the other, the stronger will be the latter's sentiment of respect (or hostility) toward him, and the more nearly will the frequency of interaction be kept to the amount characteristic of the external system*" (p. 247, his italics). Yet it is quite

clear in his discussion here and elsewhere (cf. esp. pp. 182, 418) that what counts is not *who* originates the interaction, but the *nature* of the interaction originated. A leader or person in authority is identified by the *kind* of interaction he receives *and* originates, not by the gross amount originated and received. In other words, interaction is too gross a concept for the demands Homans places on it. Similar strictures will occur to most psychologists about *sentiment*, which embraces the entire area of drives, emotions, feelings, and attitudes.

While Homans makes no reference to recent experimental studies of group dynamics, many of his hypotheses are experimentally researchable and should stimulate productive investigation. And the good use he makes of naturalistic observation should remind psychologists of the potential value of nonexperimental approaches in defining profitable hypotheses in this new field.

M. BREWSTER SMITH.

Vassar College.

KUBIE, LAWRENCE S. *Practical and theoretical aspects of psychoanalysis*. New York: International Universities Press, 1950. Pp. xvii+252. \$4.00.

In this expanded revision of his book, *Practical Aspects of Psychoanalysis*, first published in 1936, the author addresses himself to prospective patients and to persons concerned with them as well as to physicians and students of psychiatry and psychology. He hopes to explain what psychoanalysis is, clarify misunderstandings, aid in the choice of an analyst, and answer the many questions which are commonly asked about psychoanalysis.

Changes made in this edition reflect alterations in attitudes toward psychoanalysis and the greater sophistication of potential readers which have come about in the fifteen intervening years. There is, as the author says, "less occasion for indignation" because even the misunderstandings which persist are on a more reasonable level. Blind hostility has largely abated. Analysts, moreover, have learned the necessity for variety in theory and practice and for freedom for research. The emphasis has therefore shifted to accommodate the present more informed audience.

Psychoanalysis is defined as "a specific technique for studying and influencing the form, the distribution, and the utilization of psychological forces." Its goal is to broaden the domain of conscious control in human life by uncovering and modifying unconscious psychological forces. In explaining how this is done, considerable exposition of the theoretical structure of psychoanalysis is undertaken. There are discussions of the analysis of the transference, the role of dream analysis, and the technique of free association. In keeping with the psycho-

analytic conception of a universal neurotic process, the statistical idea of normality is abandoned. Psychological "normality"—which might better be called "optimal health"—is defined as determination by conscious rather than unconscious forces. Although the necessary theoretical exposition forms an important part of the book, the greatest emphasis is placed upon practical considerations.

Much of the confusion about psychoanalysis arises from the relationship between the analyst and his patient. This is, therefore, discussed at length and in terms which are designed to clear away any suggestion of mystery. The necessity for the "analytic incognito" is made clear, and the transference is explained. The ethical code of the psychoanalysis is stressed. The relationships of the referring physicians, relatives, and friends are gone into thoroughly.

Many readers will be interested in the chapter dealing with cost of psychoanalysis. A psychoanalyst, unlike other medical specialists, must make his entire income from a very limited number of patients and must set his fees according to what these patients can afford. The average fee of \$14.50 per session may not produce an unduly high yearly income for the psychoanalyst, but when this fee is multiplied by several weekly sessions over a period of a year, it is higher than the average annual income of a large proportion of families. Clearly, psychoanalysis is an expensive undertaking and one that must be confined, for the most part, to well-heeled citizens. The author is aware of this problem and suggests that the community as a whole is responsible for its solution.

One of the charges commonly leveled against psychoanalysis is that it frequently results in divorce. Dr. Kubie insists that this is an exaggerated impression, and gives cogent reasons. There are also chapters on the contrast between psychoanalysis and faith-healing, and on psychoanalysis and moral responsibility.

In general the author hews close to the line of conservative Freudian theory, deploring such modifications in technique as the presently popular reduction in number of weekly sessions, and such theoretical developments as the emphasis upon cultural factors in the development of neurosis.

In his chapter, "Controversies and Frontiers," he points out the areas in which technical, theoretical, and clinical research are needed. This section will appeal more to his fellow-workers in the field than to the general public, and it should be of interest to students looking for a problem in this field. Professional persons will also be interested in his comments on training matters and on the need for nonmedical psychotherapists.

Dr. Kubie feels that psychoanalysis is the answer to many of the world's most serious problems, and he argues effectively enough for his point of view that he does not seem to be fatuously sponsoring a panacea. He is certainly qualified to define and elucidate his specialty, and

the book should be of value to the wide audience to which it is addressed.

JESSIE L. MILLER AND JAMES G. MILLER.

University of Chicago.

FISHER, V. E. *The meaning and practice of psychotherapy*. New York: Macmillan, 1950. Pp. xv+411. \$5.00.

This is a book of transition, marking a twilight zone between two eras of psychiatric thought. Twenty-five years ago, textbooks on abnormal psychology and psychopathology were still almost entirely *descriptive*—categorical, organically minded, and fatalistic. Fisher's volume points up the contrasting outlook which has come with the newer emphasis upon *psychodynamics* and the possibilities thus opened up for better understanding and more effective treatment.

But the book has a heavy and debilitating imprint of the past upon it. After a brief and fumbling account of what the author considers to be the principles of psychotherapy, which occupies a single chapter of 44 pages, the rest of the book (22 chapters) is devoted to a succession of case histories which are certainly more reminiscent of Kraepelin than of Freud. True, there is throughout a nominal emphasis upon psychotherapy, but it is therapy which is ordered by diagnostic classification rather than by dynamic principles that cut across classifications. For each of the classical syndromes a highly specific type of treatment is described and recommended, as if one were dealing with so many definite and distinct disease entities, for which rational and relatively independent remedies had been discovered.

Many of the clinical portraits are well drawn, and shrewd insights appear on occasion. But in general this book is conceptually so uncoordinated that it will serve as a satisfactory guide to practice only for those who are content to work by rote and prescription rather than within the framework of a consistent, systematic theory of human personality and its vagaries.

O. H. MOWRER.

University of Illinois.

DEMING, W. E. *Some theory of sampling*. New York: Wiley, 1950. Pp. xvii+602. \$9.00.

Book-publishing in the field of statistics, both theoretical and applied, is undergoing a great and continuing boom. Although there is a certain amount of chaff in this production, in the main the volumes are remarkable for the high level of quality maintained. This high quality level can be illustrated by mention of just a few of the fine publications appearing in the past two years: Cochran and Cox, *Experimental De-*

signs; Mood, *Introduction to the Theory of Statistics*; Wald, *Statistical Decision Functions*; Yates, *Sampling Methods for Censuses and Surveys*. As far as the writer knows, the above-mentioned work of Yates was the first publication in book form dealing rather exclusively with survey sampling methods. The book under review is the second; it too continues the high-quality level.

As the title indicates, all the material in this book has to do with sampling. However, treatment of one aspect of sampling—namely survey methods—dwarfs treatment of every other aspect. The survey material might make about 350 pages, a respectable book on its own, and the miscellaneous mathematical topics might make another book—each worth while in its own right.

This work is lengthy and, what with the extensive use of fine type, has even more material than its 600 pages suggest. First, in fine type there are dozens of "Remarks," little helpful asides to the reader telling extra details: just what form of an equation is best for practical work, special cost features of a method, exceptional cases, other examples illustrating use of a method, and so forth. Second, there are many worked-out exercises and problems illustrating in considerable detail the application of the methods discussed. Third, mathematical derivations are shown in great detail with few steps omitted, and frequently alternative derivations of the same result are given, offering the reader that little boost often so helpful. The reviewer hopes this extra care in derivations will not backfire to discourage readers, because such care increases the ratio of equations to text, thereby suggesting need for a higher degree of mathematical sophistication than is actually required for reading the survey material. On the other hand, many texts by slipping in an equation only occasionally and omitting all important steps give an illusion of simplicity of which more wary readers may by now be quite suspicious. To give an idea of the detail—in the discussion of ratio-estimates ten examples of possible uses are given, written out completely, including notation. This thoroughness is emphasized to give an idea of the reference value of the book. The figures look professional, and the equations are as good-looking as a printer can make them. As the author points out, certain topics such as optimum allocation in stratified multistage sampling are barely mentioned. But, on the whole, the coverage is very wide.

After reading this volume, many are going to feel that the good old days of "let's send out a questionnaire," or "let's have a survey" are just about over. The faults of surveys are beginning to be quite well known, and the ways of adjusting these faults are not usually trivial. It begins to appear that an investigator armed merely with a clear conscience and \sqrt{npq} is not adequately equipped for planning, executing, or interpreting factual, opinion, or attitudinal surveys.

Psychologists interested in surveys are bound to be interested in this

treatment, which in many ways is more specific and detailed than that of Yates.

FREDERICK MOSTELLER.

Harvard University.

FRYER, D. H., & HENRY, E. R. (Eds.) *Handbook of applied psychology*. (2 vols.) New York: Rinehart, 1950. Pp. xix+382; ix+383-842. \$12.50.

In this work the editors have brought together 115 brief essays, averaging about six pages each in length, written by over one hundred contributors and dealing with all aspects of applied psychology. The attempt has been to be all-inclusive, in the sense that articles have been prepared relating to each applied field with which psychologists concern themselves. The attempt has been to be inclusive also in that the editors have tried to provide material on the techniques which are used, the typical patterns of administrative organization, and the desirable training for particular fields, as well as the results of investigation on special topics and problems.

To spread out over this range the book has necessarily had to be spread thin. Thinness is apparent in part in the amount of space allotted to a given topic. Thus, there is a five-page article on "Child Development," a seven-page article on "Aptitude and Intelligence," a six-page article on "Guided Learning," and the like. It has typically been necessary for the treatment of a topic to be either very skeletonized or else very selective, presenting only one or two aspects of the topic. As an illustration of the first, one might mention an article on "Speech Difficulties," where within nine pages of text the author sketches in the characteristics and the therapy for some thirty categories of speech disorder. The second approach is seen in an article on "Product Testing," in which the author limits himself largely to five illustrations of surveys with which he had been personally familiar.

In part, the thinness is apparent in gaps in the areas dealt with. The articles represent shafts sunk here and there, rather than a uniform level of exploitation of the different areas. Thus, in the chapter dealing with techniques of personnel psychology, there is a treatment of validity but no discussion of reliability, an article on item analysis but no consideration of item-writing. There is a section on child development but no treatment of adolescence, adulthood, or old age. Other examples could be cited.

This book cannot, therefore, be considered a fundamental treatise on applied psychology or any section of it. No single topic is treated with a fullness which makes it a definitive treatment, as has been true of some of the past handbooks which have appeared in experimental psychology, social psychology, or child development. In fact, it does not really seem that this was the editors' purpose. The book consists rather of a series

of tastes. But who is supposed to be the taster? Is it the beginning student who is taking an introductory survey course in applied psychology? The advanced undergraduate who is trying to decide whether he is interested in psychology as a life occupation? The graduate student who is trying to decide just what form of specialization within psychology he shall follow? It is hard to tell. The variation from article to article is such that the audience seems to be now one, now the other. A three-page discussion of "Drugs and Smoking" would have little value above an introductory course, but not far from this we find several articles which seem to presuppose familiarity with the concepts of factor analysis.

The reviewer was most interested in the latter parts of this work in which an attempt was made to describe the work done by the psychologist in different settings and to outline the training needed for different types of psychological service. Bringing together material of this type seemed to be the most original and distinctive contribution of the document. Some of the articles which gave fairly full discussions of one or two enterprises in which the author had been engaged seemed also to be of interest and value. Articles which provided brief and necessarily incomplete summaries of the research work on a given problem seemed likely to add relatively little to a good text. In general, the reviewer feels that for undergraduate students the book would have been more satisfactory if it had provided more by way of concrete and specific illustrations of the work the applied psychologist does, and less discussion at the level of generalities and of condensed summarization of research work. The beginning student could well profit from more intimate contact with the psychologist at work. The advanced student who wants to be fully informed on a given topic will need to go to more specialized sources than the *Handbook of Applied Psychology*.

ROBERT L. THORNDIKE.

Teachers College, Columbia University.

SCHEINFELD, AMRAM. *The new you and heredity*. Philadelphia: Lippincott, 1950. Pp. xxii+616. \$5.00.

The first edition of Scheinfeld's *You and Heredity*, published in 1939, won immediate recognition as an outstanding achievement in the writing of popular science. It was selected for distribution to members of the Book-of-the-Month Club and received high praise from scientists in the fields of biology, anthropology, sociology and medicine. The new volume, however, is far better; about 80 per cent of the earlier book has been completely rewritten, and new material to the extent of some two hundred pages has been added. In fact the amount of information here brought together is remarkable, and all of it is presented in such a way as to make it seem important to the reader personally. The result is a degree of reader-appeal that is rarely equalled by a nonfiction book.

One who picks it up for a cursory examination is more than likely to read on and on for hours, and then to lay it down reluctantly.

The book contains forty-eight chapters. The first ten present briefly and with admirable clarity the basic mechanisms of inheritance. Then follow twenty-one chapters which deal with the genetics of eye color, hair color, skin color, body build, multiple births, longevity, physical abnormalities, mental disorders, feeble-mindedness, blood types, and specific disease tendencies. Chapter 31, under the title "'Black Gene' Roll-Call," lists more than one hundred fifty kinds of physical and mental abnormalities together with a brief statement of what is or is not known about the genetic factors involved in each. The chapter closes with "forecast tables" for predicting the probable results of matings involving recessive genes, sex-linked recessive genes, dominant genes, and sex-linked dominant genes. This chapter, although lacking the dramatic qualities which characterize so many of the others, is one of the most valuable in the book.

Thus the first thirty-one chapters are devoted almost wholly to the factual results of genetic research. The remaining chapters (which might well have been called Part II) deal with issues that are much more controversial, including, among others, mental tests and their interpretation, special talents, genius, behavior traits, personality and temperament, sexual behavior, criminality, race differences, fertility trends, and eugenics. Because of the wide range of subject matter in these later chapters it was of course impossible for the author to review the available literature thoroughly on any topic. To have done so would have required a volume each for the genetics of racial and class differences, sibling and twin resemblances, special talents and genius, eugenics, etc. It was a choice between treating these controversial issues briefly and summarily or omitting them altogether. In our opinion the author's choice of the former alternative was wise. Precisely because so many of the nature-nurture issues *are* controversial and emotion-provoking, it is all the more important that the general reader should be given some orientation in regard to what is known to be true, what is probably true, what is pretty certainly untrue, and what is still in the realm of speculation.

This is what the author has tried to accomplish, and in the attempt he has maintained a commendable degree of objectivity. Since nature-nurture investigators do not agree among themselves they can be expected to react differently to some of Scheinfeld's interpretations. There are psychologists who will think the author has underestimated race differences in native abilities, but few if any will disagree with his conclusion that even if such inherent differences do exist between race averages, they are overshadowed by the individual differences within a given race. As for IQ differences between the social classes, much the same could be said except that the evidence for genetic factors is here

more abundant and less ambiguous. All but the most die-hard environmentalists would agree with Scheinfeld in preferring for adoption the child of gifted rather than dull parentage. We suspect that this preference of the author would have been strengthened by a more critical survey of the literature on adopted children. For example, the two most carefully controlled studies of adopted children—those of Barbara Burks at Stanford and Alice Leahy at Minnesota—are not mentioned; the only evidence referred to in this connection is that provided by the highly questionable studies of Skodak and Skeels at the University of Iowa. However, in his discussion of twin resemblance and of general “educability” the author takes his stand definitely with the moderate hereditarians.

In so wide a coverage of topics, on many of which the data available are far from complete, it would be too much to expect that every tentative conclusion or interpretation of the author will be confirmed by later research. As the author himself points out over and over, genetics is still an infant science. The important thing is that few writers of popular science have ever gone to so much trouble as has Scheinfeld to avoid errors both of fact and of interpretation. He has sought and obtained the help of one or more qualified geneticists on practically every topic treated, especially those dealt with in the first thirty-one chapters. Such wholehearted assistance of the ablest scientists can only be enlisted by an author whose “scientific conscience” is beyond question.

In short, we regard this book as a masterpiece of its kind. If there is a better example of science-for-the-layman we do not know what it is; nor do we know of any book either in the sciences or the humanities that more deserves to be read by every college student, whatever his major field of interest.

LEWIS M. TERMAN.

Stanford University.

GULLIKSEN, HAROLD. *Theory of mental tests*. New York: Wiley, 1950. Pp. xix+486. \$6.00.

This is an important book—a must for a large number of psychologists. It is not just another book. It is a book that fills a void by bringing together in a unified fashion the scattered literature and new material on test theory. Its author is not only qualified to do an expert job—he did an expert job.

Gulliksen starts off with four chapters on the concept of measurement errors. Fundamental relationships are derived first from a definition of error and then from a definition of true score. Reliability is defined in terms of parallel tests in such a way as to avoid circular reasoning. The exposition on reliability is straightforward, and the reader with only a modicum of elementary statistics will have no trouble

following the argument until he encounters, in chapter 5, the explanation of error as interaction. Here a minor slip occurs: whereas the s^2 values prior to page 50 are simple sample variances, the s^2 values in the development on pages 50-57 are unbiased estimates of variance; yet the two types are used interchangeably. Correction for this will lead to minor changes in equations 16, 26, and 39, and in the verbal statements based thereon.

The next three chapters are devoted to the effect of test length on reliability and validity. These chapters are clearly written and include some instructive diagrams. The Thomases who doubt "prophecy" formulas, such as the Spearman-Brown, should read pages 64-67.

In chapters 10-13 consideration is given to the effect of group heterogeneity on reliability and validity, with correction formulas for various types of selection on the ability being tested, on the criterion, and on associated variables. Chapter 13 will be omitted by the reader unfamiliar with matrix algebra, and it is possible that those who can read this chapter, though admiring the neat presentation, may raise the question as to how many ever encounter situations where correction for multivariate selection is feasible. About half of chapter 10 involves an analysis (following Mollenkopf) which purports to show that whether or not measurement errors are correlated with test score depends upon the skewness and kurtosis of the distribution of test scores. The reviewer is highly skeptical of certain parts of this derivation and equally skeptical of the outcome. For instance, the end result simply does not consistently predict the known direction of the relationship between measurement errors and test scores on the 1937 Stanford-Binet.

In chapter 14 one finds the development of a statistical criterion for judging whether tests are parallel, i.e., meet the requirements of equality of means, equality of variances, and equality of intercorrelations (the last in case of three or more forms). The question of the experimental methods of obtaining test reliability is next considered (chap. 15). This is an excellent discussion of the nonstatistical aspect of reliability determination. Next there is a chapter on estimating reliability by way of item homogeneity, and a chapter on speed versus power tests.

Chapter 18, on methods of test scoring (usually "the number of items answered correctly is an eminently satisfactory score"), is followed by a fifty-page chapter on methods of standardizing and equating scores. Here one finds a first-rate discussion of the various scores derived by transformations. The author points out the difficulties encountered in age scales, particularly those difficulties which arise from choice of the regression line used in defining mental age. He does not however, contribute anything as to which regression line is the worse for this purpose. The author has nothing favorable to say for the

MA-IQ scheme of scoring. A more objective account would have listed both the advantages and disadvantages of such scoring.

Chapter 20 is on the problem of weighting and differential prediction. It is first shown that little is gained by weighting items unless the number of items is small. Weighting for optimum prediction, weighting according to reliability, weighting inversely as the standard deviation, weighting by judgment and by factor methods, and weighting to maximize reliability, are all discussed. The dangers involved in using so-called differential tests, leading to patterns or profiles, are pointedly set forth; it is to be hoped that our clinical friends note his discussion.

The final chapter, which is concerned with item analysis, was somewhat of a disappointment to the reviewer, who had hoped to find therein a critique of the many and sundry schemes which are currently advocated and used in item work. Instead, one finds a noteworthy consideration of the important topic of item parameters as related to the mean, variance, reliability, and validity of a test. This sets the stage for the type of item analysis advocated by Gulliksen, and which the reviewer heartily supports. Simply determine for each item its difficulty, its standard deviation, its "reliability index" (defined as its point-biserial correlation with total score multiplied by its *SD*), and its "validity index" (its point-biserial with the criterion multiplied by its *SD*).

Each chapter ends with a neat summary, followed by a set of exercises. A bibliography of some five hundred titles is included. The general typographic style is pleasing, and the text is practically free of the nuisance errors which are so difficult to detect in mathematical material.

Gulliksen has done an excellent, highly commendable job. The few critical points raised by the reviewer are indeed minor in contrast with the innumerable instances deserving praise. This volume represents a valuable contribution; it is a volume of distinction.

QUINN MCNEMAR.

Stanford University.

BOOKS AND MATERIALS RECEIVED

ABT, LAWRENCE EDWIN. (Ed.) *Human engineering*. Ann. N. Y. Acad. Sci., 1951, 51, 1123-1278.

AMERICAN MISSIONARY ASSOCIATION. RACE RELATIONS DEPARTMENT. *Human rights and human relations. A report of discussions of the Sixth Annual Institute of Race Relations*. Nashville, Tenn.: Fisk Univ., 1951. Pp. 92. \$2.00.

BABCOCK, HARRIET. *The MacQuarrie Test as a clinical instrument*. Lancaster, Pa.: Science Press, 1950. Pp. 72. \$2.00.

BERGLER, EDMUND. *Neurotic counterfeit-sex*. New York: Grune & Stratton, 1951. Pp. xii+360. \$5.50.

BERRIEN, F. K. *Comments and cases on human relations*. New York: Harper, 1951. Pp. xi+500. \$4.50.

BLAKE, ROBERT R., AND RAMSEY, GLENN V. *Perception: An approach to personality*. New York: Ronald, 1951. Pp. viii+442. \$6.00.

CLARK, WALTER HOUSTON. *The Oxford Group: Its history and significance*. New York: Bookman Associates, 1951. Pp. 268. \$3.50.

COBHAM, E. M. *Mary Everest Boole: A memoir with some letters*. Rochford & Essex, Eng.: C. W. Daniel, 1951. Pp. xiv+144. 10s. 6d.

COLBY, KENNETH MARK. *A primer for psychotherapists*. New York: Ronald, 1951. Pp. viii+167. \$3.00.

DEUTSCH, MORTON, AND COLLINS, MARY EVANS. *Interracial housing: A psychological evaluation of a social experiment*. Minneapolis: Univ. of Minnesota Press, 1951. Pp. xv+173. \$3.00.

EDUCATIONAL POLICIES COMMISSION. *Moral and spiritual values in the public schools*. Washington, D. C.: National Education Assn. of the U. S., 1951. Pp. x+100. \$1.00.

ELLIS, ALBERT. *The folklore of sex*. New York: Charles Boni, 1951. Pp. 313. \$5.00.

ELLIS, ROBERT S. *Educational psychology: A problem approach*. New York: Van Nostrand, 1951. Pp. xi+546. \$5.00.

FORD, CLELLAN S., AND BEACH, FRANK A. *Patterns of sexual behavior*. New York: Harper & Paul B. Hoeber, 1951. Pp. viii+307. \$4.50.

FROELICH, CLIFFORD P. *Guidance workers' qualifications: A review of the literature, 1947-1950*. Washington, D. C.: Federal Security Agency, Office of Education, 1951. (Misc. 3376). Pp. 16.

FROMME, ALLAN. *The psychologist looks at sex and marriage*. New York: Prentice-Hall, 1950. Pp. xv+248. \$2.25.

GOHEEN, HOWARD W., AND KAVRUCK, SAMUEL. *Selected references*

on test construction, mental test theory, and statistics, 1929-1949. Washington, D. C.: U. S. Civil Service Commission, 1950. Pp. xii+209. \$1.50.

HARMS, ERNEST. *The handbook of child guidance*. New York: Child Care Publications, 1947 (1950). Pp. 751. \$6.00.

HATHAWAY, STARKE R., AND MEEHL, PAUL E. *An atlas for the clinical use of the MMPI*. Minneapolis: Univ. of Minnesota Press, 1951. Pp. xlv+799. \$9.75.

HELSON, HARRY. (Ed.) *Theoretical foundations of psychology*. New York: Van Nostrand, 1951. Pp. xix+787. \$7.25.

HORROCKS, JOHN E. *The psychology of adolescence*. Boston: Houghton Mifflin, 1951. Pp. xxvi+614. \$4.50.

KITZINGER, HELEN, AND BLUMBERG, EUGENE. *Supplementary guide for administering and scoring the Wechsler-Bellevue Intelligence Scale (Form I)*. *Psychological Monographs* No. 319 (Vol. 65, No. 2). Washington, D. C.: American Psychological Assn., 1951. Pp. v+20. \$.75.

KUDER, G. FREDERIC, AND PAULSON, BLANCHE B. *Exploring children's interests*. Chicago: Science Research Associates, 1951. Pp. 48. \$.40.

LEWIN, KURT. *Field theory in social science: Selected theoretical papers*. New York: Harper, 1951. Pp. xx+346. \$5.00.

LINDQUIST, E. F. (Ed.) *Educational measurement*. Washington, D. C.: American Council on Education, 1951. Pp. xix+819. \$6.00.

LUCAS, DARRELL BLAINE, AND BRITT, STEUART HENDERSON. *Advertising psychology and research*. New York: McGraw-Hill, 1950. Pp. xi+765. \$6.50.

MCMENAR, QUINN. *Exercises and tables of squares and square roots for "Psychological Statistics"*. New York: Wiley, 1951. Pp. 345-380. Gratis.

MASLOW, A. H., AND MITTLEMANN, BELA. *Principles of abnormal psychology* (Rev. Ed.). New York: Harper, 1951. Pp. xv+665. \$5.00.

MEYER, ADOLF. *The collected papers of Adolf Meyer*. Vol. I, *Neurology*; Vol. II, *Psychiatry*. Baltimore: Johns Hopkins Press, 1950. Pp. xxiv+693; xx+674. \$30.00, 4-vol. set.

MILLER, DELBERT C., AND FORM, WILLIAM H., *Industrial sociology: An introduction to the sociology of work relations*. New York: Harper, 1951. Pp. xi+896. \$6.00.

MUNN, NORMAN L. *Psychology: The fundamentals of human adjustment*. (2nd Ed.) Boston: Houghton Mifflin, 1951. Pp. xvi+624. \$4.50.

MURPHY, GARDNER. *An introduction to psychology*. New York: Harper, 1951. Pp. xvii+583. \$4.25.

MYERS, GARRY C., AND MYERS, CAROLINE CLARK. *Homes build persons*. Philadelphia: Dorrance, 1950. Pp. v+329. \$3.00.

OGLE, KENNETH N. *Researches in binocular vision*. Philadelphia, Pa.: Saunders, 1950. Pp. x+345. \$7.50.

PERMANENT INTERNATIONAL COMMITTEE OF LINGUISTS. *Linguistic bibliography for the years 1939-1947*. Vols. I & II. Utrecht-Brussels: Spectrum Publishers, 1949. Pp. xxiv+589.

PIÉRON, HENRI FESSARD, ALFRED, AND FRAISSE, PAUL. *L'année psychologique*. (Quarante-neuvième année, 1948.) Paris: Presses Universitaires de France, 1948. Pp. vii+700. 1,800 frs.

PRONKO, N. H., AND BOWLES, J. W., JR., *Empirical foundations of psychology*. New York: Rinehart, 1951. Pp. xvi+464. \$4.00.

RAMSEY, GLENN V. *Factors in the sex life of 291 boys*. Madison, N. J.: Glenn V. Ramsey, 82 Samson Ave., 1950. Pp. vii+106. \$2.00.

RAPP, ALBERT. *The origins of wit and humor*. New York: Dutton, 1951. Pp. 208. \$2.75.

REDL, FRITZ, AND WATTENBERG, WILLIAM W. *Mental hygiene in teaching*. New York: Harcourt, Brace, 1951. Pp. xiii+454. \$5.00.

REMMERS, H. H., DRUCKER, A. J., MYERS, M. S., AND BENNETT, E. M. *Some personality aspects and religious values of high school youth*. Publ. Opin. Panel, Vol. 10, No. 3. Lafayette, Ind.: Purdue Univ., Division of Educational Reference, 1951. Pp. 8+21a. Mimeo.

REMMERS, H. H., AND HACKETT, C. G. *What are your problems?* Chicago: Science Research Associates, 1951. Pp. 48. \$40.

ROBBINS, SAMUEL D. (Comp.) *A dictionary of speech pathology and therapy*. Cambridge, Mass.: Sci-Art, 1951. Pp. 112. \$1.65.

ROHRER, JOHN H., AND SHERIF, MUZAHER. *Social psychology at the crossroads: The University of Oklahoma lectures in social psychology*. New York: Harper, 1951. Pp. viii+437. \$4.00.

ROSCOE, STANLEY N. *Flight by periscope*. Univ. of Ill. Bull. No. 48., Aeronautics Bull. No. 9. Urbana, Ill.: Univ. of Illinois, 1951. Pp. 46.

SEWARD, GEORGENE H., MORRISON, LESTER M., AND FEST, BEVERLY. *Personality structure in a common form of colitis*. *Psychological Monographs* No. 318 (Vol. 65, No. 1). Washington, D. C.: American Psychological Assn., 1951. Pp. v+26. \$1.00.

SOCIETY FOR EXPERIMENTAL BIOLOGY. *Symposia. IV. Physiological mechanisms in animal behavior*. New York: Academic Press, 1950. Pp. vi+482. \$6.00.

SYMONDS, PERCIVAL M. *The ego and the self*. New York: Appleton-Century-Crofts, 1951. Pp. ix+229. \$2.50.

THIEL, MANFRED. *Versuch einer Ontologie der Persönlichkeit*. Berlin: Springer, 1950. Pp. xxiii+635. DM 29.70.

THURSTONE, THELMA GWINN, AND BYRNE, KATHARINE MANN. *Mental abilities of children*. Chicago: Science Research Associates, 1951. Pp. 48. \$40.

TOLMAN, EDWARD CHACE. *Collected papers in psychology*. Berkeley & Los Angeles: Univ. of California Press, 1951. Pp. xiv+269. \$4.50.

TOMAN, WALTER, *Einführung in die moderne Psychologie*. Vienna: Humboldt, 1951. Pp. 268.

TRNKA, B. *A tentative bibliography [of writings on linguistic statistics]*. A publication of the Committee on Linguistic Statistics, of the Permanent International Committee of Linguists. Utrecht-Brussels: Spectrum Publishers, 1950. Pp. 22.

VERNON, PHILIP E. *The structure of human abilities*. New York: Wiley, 1950. Pp. 160. \$2.75.

WALLON, HENRI, AND EVART-CHMIELNISKI, EUGÉNIE. *Les mécanismes de la mémoire*. Paris: Presses Universitaires de France, 1951. Pp. viii+116. 400 frs.

WAY, LEWIS. *Adler's place in psychology*. New York: Macmillan, 1950. Pp. 334. \$4.50.

WEPMAN, JOSEPH M. *Recovery from aphasia*. New York: Ronald, 1951. Pp. xii+276. \$4.50.

WOODRUFF, ASAHEL D. *The psychology of teaching*. (3rd Ed.) New York: Longmans, Green, 1951. Pp. xix+617. \$4.75.

WRENN, C. GILBERT. *Student personnel work in college*. New York: Ronald, 1951. Pp. ix+589. \$4.75.

YACORZYSKI, G. K. *Medical psychology: A basis for psychiatry and clinical psychology*. New York: Ronald, 1951. Pp. xviii+535. \$6.00.

YOUNG, J. Z. *The life of vertebrates*. London, New York: Oxford Univ. Press, 1950. Pp. xv+767. \$8.50.

:
.
.
y
0.
:
].
r-
s:
k:
n-
p.
n,
ld,
ew
k:
nd
ord